HIGH RESOLUTION ENSEMBLE ERROR GROWTH AND DIMENSIONALITY IN TROPICAL CYCLONE GENESIS ENVIRONMENTS

by

Levi Sterling Thatcher

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Atmospheric Sciences

The University of Utah

August 2013

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of Levi Sterling Thatcher

has been approved by the following supervisory committee members:

Zhaoxia Pu	, Chair	5/6/2013 Date Approved
John D. Horel	, Member	5/6/2013 Date Approved
Steven K. Krueger	, Member	5/6/2013 Date Approved
Carolyn Reynolds	, Member	5/6/2013 Date Approved
Jeffrey Anderson	, Member	5/6/2013 Date Approved

and by Kevin D. Perry , Chair of

the Department of Atmospheric Sciences

and by Donna M. White, Interim Dean of The Graduate School.

ABSTRACT

Over the last several decades, ensemble forecasts of atmospheric phenomena have become increasingly popular, not only because they provide an improved mean forecast of various events, but also because they render an estimate of the accompanying forecast uncertainty. Research into high-resolution ensembles based in the Tropics and in terms of tropical cyclone (TC) genesis mechanisms has been relatively sparse, even though such disturbances are notoriously difficult to forecast.

In this study, we couple several popular ensemble perturbation methods to the mesoscale Weather Research and Forecasting (WRF) model at high resolution to examine the predictability of genesis, error growth characteristics, and underdispersion issues in forecasts of Hurricane Ernesto (2006) and Typhoon Nuri (2008). In order to examine the effects of model resolution on TC genesis forecasts, a downscaled 5-km resolution regional control ensemble, based on a downscaling of the National Centers for Environmental Prediction's Global Ensemble Forecast System (GEFS), is compared against the standard GEFS simulations. To analyze the effect of the various perturbation methods on genesis and forecast characteristics, we compare results from the regional GEFS-based simulation to several implementations of the breeding of growing modes (BGM), wherein we vary the variables perturbed, cycling period durations, and boundary conditions.

While the global GEFS forecast failed to predict a well-developed Ernesto in any of its members, the high-resolution GEFS-based ensemble contained several intense TCs

by actual genesis time. Based on a sample of 154 ensemble member forecasts, the impact of environmental precursors on TC genesis likelihood is investigated. Despite the large number of easterly waves that do not develop into TCs and the large amount of water vapor in the summer Tropics, we find that the strength of the preexisting wave and initial 850 hPa water vapor are significant determining factors for TC genesis.

Finally, we create several ensemble forecasts of Ernesto using the stochastic kinetic-energy backscatter scheme (SKEBS) and find that the standard SKEBS ensemble has more dispersion per unit error compared with both the BGM and GEFS-based ensembles. In addition, SKEBS shows notably lower vapor bias and larger theta bias compared with the initial condition-based ensembles.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Zhaoxia Pu for her guidance, encouragement, and technical help with this dissertation and the related research. The work of the committee members, Drs. John Horel, Steven Krueger, Jeffrey Anderson, and Carolyn Reynolds in improving and refining the various themes of this project has also been greatly appreciated.

Additionally, I thank my wife, Emma, for her tireless support, wit, and strength, as well as my whole family, department colleagues, and friends for their talents and wisdom.

The Center for High Performance Computing (CHPC) at the University of Utah has provided comprehensive computational support for this project and their services have been greatly appreciated. This research is supported by the ONR TCS-08 program.

CHAPTER 1

INTRODUCTION

Background

Due to the large uncertainties inherent in atmospheric observations, the initial conditions of forecast models inevitably contain errors. When making simulations based on such initial conditions, these errors are often accounted for in the creation of several simulations based on slightly perturbed initial conditions. Such ensembles of simulations not only provide a better estimate of the true state of the atmosphere compared with a deterministic forecast, but they also supply an estimate of the associated forecast uncertainty. Considering these benefits and recent increases in computing power, accurate forecasts of the atmosphere have become ever more reliant on ensemble forecasts using slightly perturbed simulations.

While atmospheric modeling began to emerge very slowly in the early 20th century, predictability and ensemble forecasts did not become formal realms of inquiry until the 1960s. Lorenz (1963) asserted that the atmosphere would always thwart forecasters' best efforts, because "two states differing by imperceptible amounts may eventually evolve into two considerably different states." This observation was made after he had repeated identical calculations using (seemingly) exact initial conditions. After being surprised by the large difference between the results, Lorenz determined that it was due to round-off error. Lorenz (1963) goes on to say, foreshadowing the new field

of chaos theory and explaining the fundamental problems addressed by the current dissertation, that "if, then, there is any error whatever in observing the present state—and in any real system such errors seem inevitable—an acceptable prediction of an instantaneous state in the distant future may well be impossible." Not long after, Epstein (1969a) summarizes the nature of the problem by wondering whether one obtains the best forecast by applying the deterministic equations to one best estimate of the initial conditions. He then further elucidated the problem many modelers currently struggle with when he states that one cannot know a uniquely valid starting point for each forecast and that there is instead an almost infinite ensemble of starting points (1969a). He goes on to make strides in solving this problem by developing what he calls the stochastic-dynamic prediction equations (Epstein 1969b). Essentially, he promoted perturbed initial states which reflect the uncertainty in the then-current knowledge of the atmosphere. These ensemble members are then treated deterministically. This ensemble approach recognizes that the state of the atmosphere can only be known in terms of probabilities and is akin to "describing a probability density function in a multidimensional phase space" (Epstein 1969b).

This brief history helps demonstrate how essential ensembles are to improve the accuracy of forecasting, which was partly evident in the fact that the necessity for these arose near the beginning of the modeling era. While the techniques have progressed and some of the questions have changed, the primary reasons for this approach remain the same. Given the sparse number of observations available over the tropical oceans, there are an innumerable number of analyses that could accurately fit what is actually known about the state of the atmosphere. Compared to those of an ensemble, deterministic

forecasts not only offer a worse forecast overall, but also a false sense of certainty. While Epstein (1969b) said that there are often an infinite number of atmospheric analysis states consistent with what is known observationally, it is often not practical to create thousands of ensemble members when simulating the atmosphere. It is important, and often sufficient, however, to create an ensemble that provides an accurate picture of the uncertainty of the forecast. This can often be done effectively with less than a score of members.

In terms of moving from Epstein's (1969b) stochastic dynamic theoretical framework to a functioning ensemble, one must work around various practical considerations, such as the resources available, and the perturbation methods that one can feasibly implement. Nevertheless, the overall principles behind the ensemble must remain intact in order to avoid many of the intrinsic problems associated with a strictly dynamical forecast. First among the theoretical tenets that must be implemented into the ensemble is the fact that this ensemble should determine the future probability distribution of the state of the flow (Descamps and Talagrand 2007).

Using what he deemed "Monte Carlo" approximations, Leith (1974) was one of the first to implement the recommendations of Epstein. In the former's work, we notice several important considerations in moving from theory to actual ensemble simulations. The most basic of these is the fact that the mean of the forecast ensemble must continue to be the estimate of the true state of the atmosphere that is best "in the least-square-error sense," and that the covariance of the ensemble serves as the simplest, best measure of the uncertainty in the ensemble of forecasts (Leith 1974). During the integrations of the ensemble members, the goal is to accurately describe the probability distribution in phase

space. As Leith (1974) mentioned, there are partial differential equations describing the evolution of this probability distribution. The problem is that, when considering the degrees of freedom in typical models, the direct integration of these equations is too computationally expensive. Thus, various perturbation methods have been developed to approximate such a calculation.

Ensemble perturbation methods

While there are many different ways by which one can attempt to span this distribution of points in phase space, there are primarily two important types of error, and thus two important types of perturbation methods. The first class of methods acknowledges that the corresponding model has deficiencies and attempts to account for this in some specific way. The second set of methods addresses uncertainties present in model initial conditions. In terms of the methods accounting for model error, a popular subclass comprises ensembles whose members are different models, which is called a mutlisystem ensemble (Krishnamurti et al. 2000; Kharin and Zwiers 2002; Hagedorn et al. 2005). Other ensembles in this model-error group choose a different physics parameterization for each member and/or use a kinetic energy backscatter technique to account for excessive model energy dissipation (Berner et al. 2011).

In contrast to most initial condition-based ensembles, which act as if the accompanying model is perfect, some model-aware perturbation methods go to the other extreme and instead perform system simulation experiments, wherein "one considers all elements of the forecasting system to be subject to error or imperfection" (Houtekamer et al. 1996). These perturbed model-based parameters encompass items such as roughness length, sea surface temperature, diffusion, physics parameterizations, gravity wave

parameterizations, albedo, and cumulus parameterization schemes (Houtekamer et al. 1996). These changes largely reflect the idea that the introduction of uncertainty into a model *should* improve the efficiency, defined as spread per unit of error, of that model. In yet another study about accounting for model deficiencies, Teixeira and Reynolds (2008) took a novel, stochastic approach in terms of subgrid-scale mixing, arguing that parameterization methods do not have to be treated as deterministic and that an analysis ensemble could be beneficially constrained by the "PDFs that are implicitly associated with a particular physical parameterization." The same authors tested an algorithm for a stochastic moist convection parameterization and found that significant ensemble spread results, in terms of several fundamental fields, from this simple scheme by itself. For similar studies, see Buizza et al. (1999) and Lin and Neelin (2002).

The other large class of schemes takes into account only analysis, or initial condition errors, and thus largely ignores the deficiencies generated by the accompanying model. Over the last 20 years, one of the most popular and simple of these schemes is the breeding method. The breeding method was introduced by Toth and Kalnay (1993) and was originally called the "breeding of growing modes," or BGM method. These same authors argue that ensemble perturbations must represent growing errors in the analysis and their perturbation construction method follows suit. A breeding cycle is started by introducing a random perturbation into a simulation, with a concurrent unperturbed simulation used as a control. After 3, 6, or 12 h of integration, the differences between the two simulations are rescaled to the size of the initial perturbation and then repositioned onto successive analysis fields (Toth and Kalnay 1993). This procedure is repeated at regular intervals until the end of the breeding period, wherein the bred errors become the

initial perturbations of a "free" simulation. This perturbation method shares much in common with how forecast errors are cycled in data assimilation (DA) schemes, in that with breeding vectors, the differential between two nonlinear forecasts is integrated forward (and scaled down at regular intervals) upon the evolving atmospheric fields (Toth and Kalnay 1997; Wei et al. 2008). A version of the method was employed in NCEP's operational system through the 1990s (Wei et al. 2008) and the technique will be one of the emphases of this study.

In terms of the more technical aspects of the BGM method, Toth and Kalnay (1993) described its growing perturbations as being very similar to those obtained as eigenmodes of L^TL "whenever the perturbations are growing fast," where $L(t_1,t_0)$ is the linear model propagator between time t_0 and t_1 and L^T is its adjoint. While in its time the breeding method did provide better results than much larger "Monte Carlo" or lagged average forecast ensembles (Toth and Kalnay 1993), later it was criticized because of the fact that its perturbation amplitude does not reflect geographical variations in observational network density (Wang and Bishop 2003). The latter authors also complained that the breeding method, unlike data assimilation schemes, does not "reduce error variance in directions corresponding to large error variance by a larger factor than error variance in directions corresponding to small error variance." Also, the method was criticized for producing perturbations whose "variance is concentrated in considerably fewer eigen-directions than there are perturbations" (McLay et al. 2008). Nevertheless, the breeding method has been particularly suited for ensemble creation partially because of the fact that it involves the dynamical cycling of ensemble perturbations (Wei et al. 2008), which well aligns with the fact that, generally, modern

NWP analysis methods strongly rely on short range forecasts (Toth and Kalnay 1993).

In the ensemble transform (ET) technique, some of the issues surrounding the breeding method's lack of interperturbation orthogonality are mitigated. The ET technique was introduced by Bishop and Toth (1999) to help decide where and when to deploy weather reconnaissance aircraft such that particularly difficult and important forecast regions could be improved. The method was later adapted for dynamically constrained sampling by Wei et al. (2004), to help reduce the eigen-spectrum issue found in the regular breeding method (McLay et al. 2008). The method proved so effective that it was implemented at NCEP starting in May of 2006 (McLay et al. 2008). Wei et al. (2008) described the ET method as a second-generation technique attempting to better link DA and ensemble creation, while McLay et al. (2008) summarized the method as constraining perturbation selection using estimates of analysis error covariance.

Another popular perturbations method is that of singular vectors (SVs), which are also called optimal vectors. These vectors are the perturbations that grow most quickly, "over a given period of time, in the dynamics linearized about a given solution of the forecast model" (Descamps and Talagrand 2007). If one first runs a tangent linear model and then the adjoint of that model (which we will call LL^T), the initial singular vectors are the eigenvectors of LL^T, a common metric whose eigenvalues are the squares of the singular values, denoted as σ_i^2 (Kalnay 2006). While bred vectors attempt to estimate fastest sustainable growth and represent probable growing analysis errors, SVs estimate vectors with fastest *transient growth* in the future (Toth and Kalnay 1997; emphasis theirs). Because of the fact that SVs are related to an integration of the adjoint model, they are more expensive than bred vectors to implement. Since the early 1990s, SVs have

been used routinely in the construction of the initial perturbations of the ECMWF ensemble prediction system (Palmer et al. 1993), and have gained a favorable reputation for their ability to accurately represent the uncertainty in a particular forecast.

Another class of perturbations methods which are affected by the error growth of the day are those that derive their perturbations from background error covariances. This perturbation technique is often accomplished by methods which parameterize or model the temporal and spatial covariance relationships of the boundary conditions (Torn et al. 2006). Within this perturbation method group, there are some techniques that use state-dependent statistics, which provide reduced boundary errors (Torn et al. 2006), and others that use assumed statistics. In terms of using assumed statistics, this could be done simply by drawing from a particular multivariate Gaussian distribution, such as the covariance model used in the Weather Research and Forecasting (WRF) 3DVAR package. In this technique, the state-independent spatial covariance used on the boundaries is derived from forecast differences and are representative of short-term forecast errors (Torn et al. 2006).

Challenges in predicting tropical cyclone genesis

Recent advancements in tropical cyclone (TC) research have led to significant improvements in track predictions (Aberson 2001), whereas making progress in terms of TC genesis and intensity change has proven exceedingly difficult (Hennon and Hobgood 2003; Rogers et al. 2006; Kerns et al. 2008). Despite a few notable studies into regional ensembles in general (Torn et al. 2006; Torn and Hakim 2008) and a few select researchers examining regional ensembles in the context of TC prediction (Torn 2010; Liu et al. 2012), a large amount of work into TCs in an ensemble context has been

performed with *global* ensembles (Mackey and Krishnamurti 2001; Cheung and Elsberry 2002; Buckingham et al. 2010; Snyder et al. 2010, 2011). The focus of this study is to fill the corresponding gaps in the literature regarding TC-based error growth characteristics and we anticipate this type of research also lends itself well to answering questions related to TC genesis mechanisms.

In the literature on TC genesis, one of the key questions is whether the surface-concentrated (warm core) tropical depression vortex results from a midlevel mesocyclone or if it has its origins closer to the ocean surface. Reasor et al. (2005) affirmed that the establishment of this surface vortex was central to the TC genesis problem. Studying the development of Hurricane Guillermo, Bister and Emanuel (1997) used an axisymmetric, nonhydrostatic model to investigate and describe the basic theory as to how a midlevel mesoscale vortex becomes a tropical depression. They show in Guillermo that a mesoscale vortex is initially found within the stratiform rain region of a Mesoscale Convective System (MCS) and that the downdrafts in this region, through the vertical advection of vorticity, helped build the surface vortex. This is typical; many studies state that cyclogenesis depends on the presence of an MCS (Chen and Frank 1993; Ritchie 1995; Simpson et al. 1997). Bister and Emanuel show that the existence of an initial cold-core vortex is crucial because of the way in which it discourages downdrafts and encourages convection. They also claim that for warm-core development to occur, stratiform rain must persist long enough such that it drives the midlevel vortex down to the boundary layer, which would take as long as it takes air to descend through the layer with evaporational cooling (Bister and Emanuel 1997). In a similar, top-down genesis vein, Ritchie and Holland (1997) describe the development of Typhoon Irving using a

variety of observational data and found that midlevel vortices play an integral role in the formation of warm core, surface-concentrated tropical depressions. Simpson et al. (1997) used a dataset similar to that of Ritchie and Holland (1997) and found that the genesis of Tropical Cyclone Oliver occurred as the circulation was built down towards the surface through the stochastic interaction of mesoscale convective vortices (MCVs) and their accompanying MCSs.

In the opposite camp are those who believe that the mechanisms associated with genesis evolve through more of a "bottom-up pathway," and are not closely related to the midlevel transport of PV typically associated with mesoscale convective systems (Montgomery et al. 2010). One of the first of such studies was that of Montgomery and Enagonio (1998), who used a nonlinear quasigeostrophic balance model to show that vortex intensification proceeds as low-level vorticity, which is generated by MCV-related convective bursts, can spin up a surface-circulation through horizontal axisymmetrization (also see Reasor et al. 2005). Similarly, Davis and Bosart (2001) examined the genesis of Hurricane Diana (1984) and found that, because the diabatic heating occurred within a frontal circulation, the important PV anomalies produced were located at low levels. Despite the fact that they only examined one case, they felt their synopsis supported a general viewpoint of mesoscale cyclogenesis. Using the MM5 at 3km resolution to analyze the genesis of the same Hurricane Diana (1984), Hendricks et al. (2004) found that the most important influence to Diana's genesis were "small-scale cores of deep cumulonimbus convection that form in a vorticity-rich environment."

Reasor et al. (2005) and Montgomery et al. (2006) subsequently helped clarify the role of cumulonimbus and vortical hot towers (VHTs) in converging vorticity and

triggering tropical cyclogenesis. While these genesis events were still explained as being dependent on the protective environment of a midlevel MCV, these were seen as bottom-up, rather than top-down ways to build the incipient TC vortex. More recently, Dunkerton et al. (2009) and Montgomery et al. (2010) relied on a more synoptic development process in proposing their marsupial pouch theory of TC genesis. In these papers, the "Kelvin cat's eye," or protective pouch, is thought to be necessary for TC genesis because it provides a region of cyclonic vorticity, weak deformation, a containment of moisture, and the maintenance of the parent (easterly) wave until the "developing proto-vortex becomes a self-sustaining entity and emerges from the wave as a tropical depression" (Montgomery et al. 2010). Throughout many of these studies, a random, or stochastic component has been described as being an integral part in the genesis process (Simpson et al. 1997). The ensemble-based focus of this dissertation will be well placed to shed light on the likelihood and nature of such phenomena. Throughout this work, multiple perturbation methods and numerous simulations will be used to determine whether the forecast TCs develop more from mid- or lower-level pre-existing vorticity.

Besides the discrepancy explained above, there are other, narrower issues that would be suitable for study with the large model output that will accompany this project. An issue ripe for examination is that of the effect of moisture on the downdrafts concomitant to TC genesis. Several recent modeling studies have looked at the effects of instability, moisture, and vertical motion on TC genesis and have found disparate results. First, Nolan et al. (2007) used a radiative-convective equilibrium model on a doubly periodic *f*-plane to investigate genesis sensitivity to environmental parameters. What they found was that there was no relationship between (moisture-dependent) CAPE and the

rate of cyclone development. This finding called into question the statistical genesis parameters of Royer et al. (1998) and DeMaria et al. (2001).

Sippel and Zhang (2008) used the MM5 in an ensemble to investigate the predictability of cyclogenesis by examining a disturbance in the Gulf of Mexico. In contradiction with Nolan et al. (2007), he found that the two most important parameters for genesis were deep moisture and CAPE. They hypothesized that higher CAPE generally favors stronger or more numerous VHTs, which have been shown to have a beneficial effect on genesis. They are careful, however, to say that CAPE is not a good predictor of genesis in general; as usual, the discussed parameters are always necessary, but not sufficient for genesis. The above findings closely tie in to the role of downdrafts in TC genesis, since if enough moisture is not present in the midtroposphere, then convection has the propensity to produce cold convective downdrafts and stabilize the lower troposphere (Sippel and Zhang 2008). Bister and Emanual (1997) similarly emphasize the importance of moisture, but hypothesize that a low-level cold core works to cause genesis by increasing the important updraft to downdraft ratio. Considering the uncertainty regarding the effects of moisture, pre-existing vorticity, and related parameters on TC genesis, the relationship between these precursors and subsequent storm formation can be carefully studied in the various regional ensembles constructed for the current project.

Scope of this study

In this study, we have chosen to study error growth that occurs in TC environments as forecast by several distinct ensemble perturbation methods coupled with a regional model because of 1) the lack of higher-resolution ensemble based studies in general; 2)

the fact that error characteristics and growth in tropical-based ensembles have seldom been studied; and 3) the disagreements still surrounding the mechanisms of TC genesis. Hurricane Ernesto (2006) is chosen for this case study because 1) of the significant damage it caused to the US and Caribbean and 2) it has been documented as being an especially difficult forecast for the NCEP GEFS ensemble (Snyder et al. 2010; Liu et al. 2012). In this study, the mesoscale regional Weather Research and Forecasting (WRF) model (Skammarock et al. 2008) is used because it represents a new-generation mesoscale model that has been widely used in research and operations worldwide.

Considering its wide usage, previous knowledge of the underlying model dynamics, competitiveness over short forecast periods (Figure 1; see also McLay et al. 2008), and ease of implementation, in this proposed study, we will examine the error growth, characteristics, spatial-spectrum, and dimensionality of the breeding method when producing forecasts of tropical cyclogenesis. When using the ET method with their NOGAPS model as compared to a simple BGM method, McLay et al. (2008) said that the former held only a modest advantage over the breeding scheme, especially over forecast periods of 0-48 hours. While the two methods do perform similarly, because of its ease of implementation, we will largely use simple variations of the BGM method to study investigate error growth, characteristics, and spatial-spectra in TC versus general tropical environments. This will be contrasted with an ET-based control ensemble as well as several implementations of a stochastic backscatter method that will represent the model-error class of perturbation methods and which will be detailed in a later chapter of the dissertation. The distinctiveness among these schemes and the fact that they have been fairly widely used over the last 10 to 20 years certainly complement the reasons

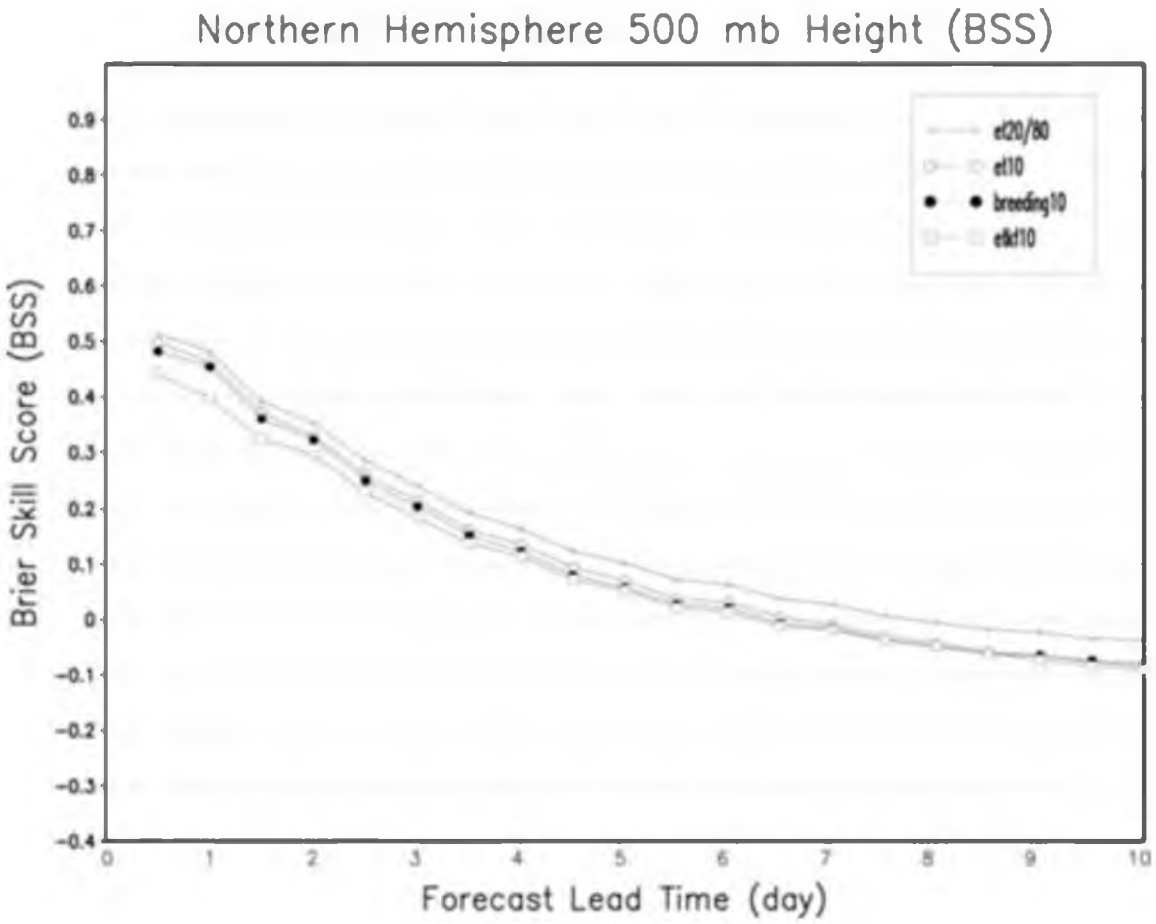

Figure 1. Averaged Brier Skill Scores of 500 mb geopotential height over the Northern Hemisphere from NCEP's breeding scheme (dark circles) and their Ensemble Transform scheme (open circles) by forecast lead time [from Wei et al. 2008, Figure 7]. (Tellus, publishes under the Creative Commons Attribution-NonCommercial 3.0 Unported license).

given above.

In addition to the uncertainty regarding which perturbation method is generally most accurate in the ensemble mean, there are also special factors related to our Tropics-based focus that we must consider. First, most perturbation methods (such as breeding vectors) are based on midlatitude variability and may not be suitable for the Tropics (Mackey and Krishnamurti 2001). Specifically, Zhang and Krishnamurti (1997) explained that perturbation growth over the midlatitudes is mainly caused by dynamical instability according to linear perturbation theory. They went on to say that the error growth mechanisms are quite different over the Tropics, and that, in particular, perturbation growth there may be more related to convection, related latent heating, and the interaction of these with the large scale. It also appears that the use of the breeding vector method over the Tropics may exacerbate the method's tendency to produce variance which is concentrated in considerably fewer eigen-directions than there are perturbations (McLay et al. 2008). Overall, seeing that TCs often provide error growth mechanisms similar to those in the midlatitudes, we can address the performance of our perturbation methods across varying error growth regimes by analyzing our results in terms of radii from the simulated TCs.

Since the relevant previous studies were largely conducted with global models, it is not known how much these tropical error growth peculiarities would be affected by the use of a relatively high-resolution regional model in the Tropics. Indeed, this general use of large-scale models has not only been the case for ensemble studies of TC genesis in particular, but also for ensemble-based studies in general. For example, Zhang and Krishnamurti (1997) used the Florida State University Global Spectral Model, Wang and

Bishop (2003) used the Community Climate Model at T42 resolution, Wei et al. (2008) used NCEP's Global Ensemble Forecast System (GEFS) ensemble, and Cheung and Elsberry (2002) as well as McLay et al. (2008) used the Navy's global NOGAPS model in their respective ensemble-based work. While there have been a few recent studies on *regional* ensemble simulations of TCs (Torn 2010; Liu et al. 2012), there has still been a significant focus on large-scale models. For example, Snyder et al. (2010) used output from NCEP's global ensemble forecast system and Snyder et al. (2011) used data from NOGAPS (also see Buckingham et al. 2010). Considering the potentially large improvements in forecast error provided by well-constructed ensembles, the gains in computing power realized over the last 20 years, and the potentially devastating impact of inaccurate TC forecasts, the need for research to examine the ability of regional high-resolution ensembles to forecast TC-related processes is abundantly clear.

In summary, with this dissertation research we aim to

- characterize error growth (in terms of structure, type, spatial spectra, and variability among members) in regional ensemble forecasts at high resolution in the vicinity of TC genesis;
- analyze the impact of perturbing different variables on ensemble error characteristics;
- study the error dimensionality differences among the various ensembles;
- investigate how well each ensemble's perturbations explain the overall forecast error (and determine how this varies spatially and temporally).

In addition, we will analyze the effect of ensemble-related logistical considerations such as the rescaling method and vertical localization. These ensemble error-related

questions will be coupled with more physical genesis-related questions addressed by the many ensembles in this study, such as

- What are the most important environmental precursors to tropical cyclone genesis?
- At what level does water vapor most promote genesis?
- Does vertical shear play a more important role than water vapor in genesis?
- Do the ensemble forecasts suggest a bottom-up or top-down route to TC genesis?

Specifically, in the first part of this work, we will examine these topics in terms of initial condition uncertainty using the breeding method. In the latter part of this work, we will use the multiphysics ensemble and stochastic kinetic-energy backscatter scheme (SKEBS; Berner et al. 2011) to further answer similar questions and strengthen our findings.

This dissertation is organized as follows: Chapter 2 describes the experimental design and the resultant genesis differences between regional and global GEFS simulations, and compares these with two simple breeding schemes; Chapter 3 discusses the differences between GEFS-based and breeding ensembles in terms of initial perturbations, error growth, and dimensionality; Chapter 4 examines the effect of cycling period and vertical localization on forecast characteristics; Chapter 5 discusses the differences between bred ensembles based on FNL versus GEFS boundary conditions as well as the impact of perturbation size on a breeding ensemble; Chapter 6 compares a multiphysics ensemble to the control and breeding ensembles in terms of various metrics; Chapter 7 discusses the physical genesis precursors as found in the ensembles; Chapter 8 analyzes the benefits of the stochastic backscatter scheme; Chapter 9 examines

similar simulations of Typhoon Nuri; and Chapter 10 provides a summary and concluding remarks.

CHAPTER 2

HURRICANE ERNESTO, ENSEMBLE AND STATISTICAL METHODS, AND BASIC EVALUATION

A brief overview of Hurricane Ernesto

Although originating as a tropical wave off the coast of Africa, Hurricane Ernesto (2006) did not achieve any notable organization until an associated surface low developed as the disturbance approached the Lesser Antilles. Moving toward the northeast, the system achieved tropical depression (TD) status 1800 UTC on 24 August 2006, roughly 40 nautical mi north-northwest of Grenada. Subsequently, the depression experienced increased convection over the low-level center as it moved north-northwest below a ridge over the western Atlantic Ocean. On 1200 UTC, the disturbance reached tropical storm (TS) status and turned to the northwest. Later, the storm reached hurricane status at 0600 UTC 27 August just south of Haiti, experienced several periods of strengthening and weakening, and eventually made landfall in Cuba, Florida, and North Carolina (National Hurricane Center Final Report).

The WRF model and setup

An advanced research version of the WRF model (version 3.3.0) (Skamarock 2008) is used with three-level nested domains and an inner-most resolution of 5 km. Detailed information on all the domain configurations is given in Table 1; domain positioning is

Table 1

Dimensions, grid spaces, and time steps for model domains in Ernesto simulations

Domain	Dimension (x × y × z)	Grid space	Time step
1	125 × 70 × 36	45 km	120 s
2	331 × 148 × 36	15 km	40 s
3	844 × 340 × 36	5 km	13.3 s

given in Figure 2. In the vertical, there are 36 σ levels and the top is set at 50 hPa. The Purdue Lin (Chen and Sun 2002) scheme is used for the microphysics, the Yonsei University scheme (Hong et al. 2006) is used for the boundary layer, and the Grell-Devenyi cumulus scheme (Grell and Devenyi 2002) is used for the cumulus parameterization, but only in the outer two domains. In addition, the Dudhia (Dudhia 1989) and the rapid radiative transfer model (Mlawer et al. 1997) schemes are used for short and longwave radiation, respectively.

The perturbation methods and experimental design

For our experiments, initial conditions (ICs) for the control ensemble (CNTL hereafter) are derived from the Global Ensemble Forecast System (GEFS) with 14 members. Because the GEFS ensemble used 14 members at that time, the rest of our ensembles will follow suit. We also use the GEFS initial perturbations to begin the breeding cycle for the ensemble experiments. Unless otherwise noted, boundary conditions (BCs) for each of our ensemble members and CNTL simulations are from NCEP's 1x1° final analysis (FNL) data, such that differences between our simulations typically exist only in terms of initial conditions.

As a control ensemble, a 14-member ensemble simulation of Ernesto is performed simply using GEFS ICs and FNL BCs, the former of which are based on the ET technique, which has been used at NCEP for the last several years (Wei et al. 2008). Next, the breeding method is constructed. This is done using the WRF model and, to our knowledge, represents one of few efforts to evaluate the breeding method in the Tropics at high resolution. To run the breeding method, it is necessary to first create a best

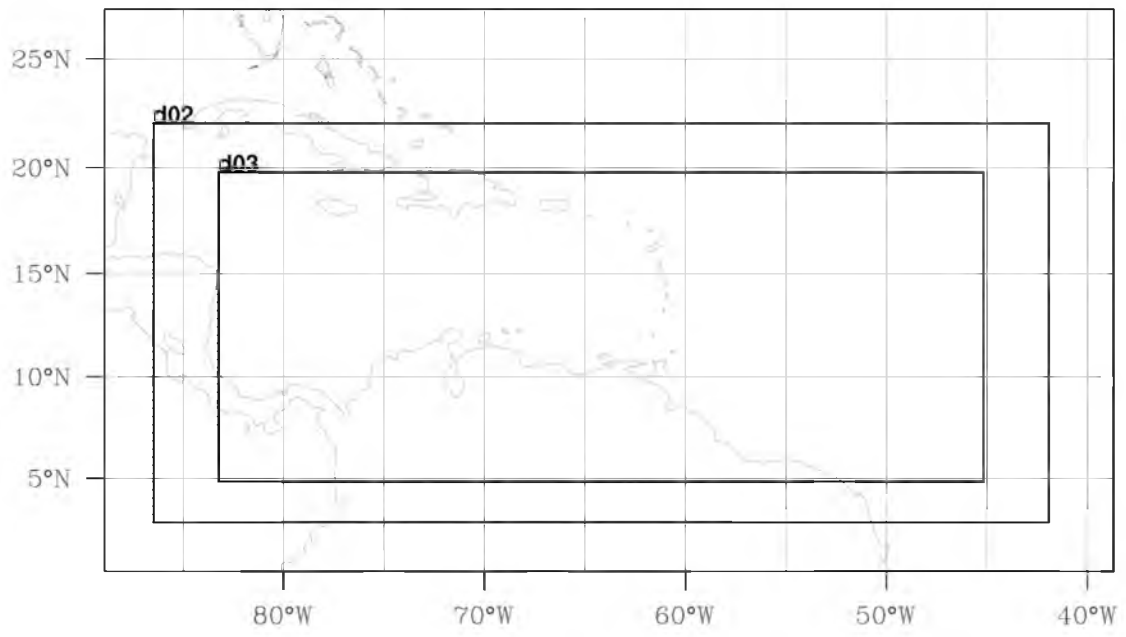

Figure 2. WRF domains used for all ensemble simulations of Hurricane Ernesto.

estimate of "truth," which we do by interpolating FNL 1x1° data down to the WRF's outer domain at 45 km resolution. We also interpolate to 5 km and later use these inner-domain analyses for ensemble evaluation. Initial fields for each of the bred members come from the corresponding GEFS member's ICs used in the control ensemble. The breeding method is subsequently constructed (see Figure 3) by performing the following for 14 members: 1) integrating the perturbed member for a predetermined cycle period at 45 km resolution; 2) differencing, in terms of the chosen state variable(s), the member's output compared to the analysis at the end of the cycle period; 3) scaling the domain- and height-averaged mean of this difference by the difference found at the beginning of the cycle; 4) adding this scaled perturbation to the analysis to create the member's initial conditions (in terms of the specified variable) for the next cycle; and 5) integrate the member for the next 6 h.

To be clear, this is a simple breeding method and no extra techniques have been used to orthogonalize the perturbations. During the 24 h breeding period, while the specified state variable fields are cycled, the other fields evolve for the 24 h without being manipulated. At the end of the 24 h period, each member's bred field replaces the corresponding field in the FNL analysis at that time, while the other fields are untouched. We then use this modified analysis (essentially updating the unperturbed fields to the current time) to start the simulation for that member (see Table 2). This setup naturally allows us to examine the effect of the individually perturbed variables on the resulting forecast. A second, similar, setup was constructed wherein the bred fields of each member were placed in the *corresponding* GEFS member's interpolated analysis at the end of the 24 h breeding period, such that here we controlled which variables were

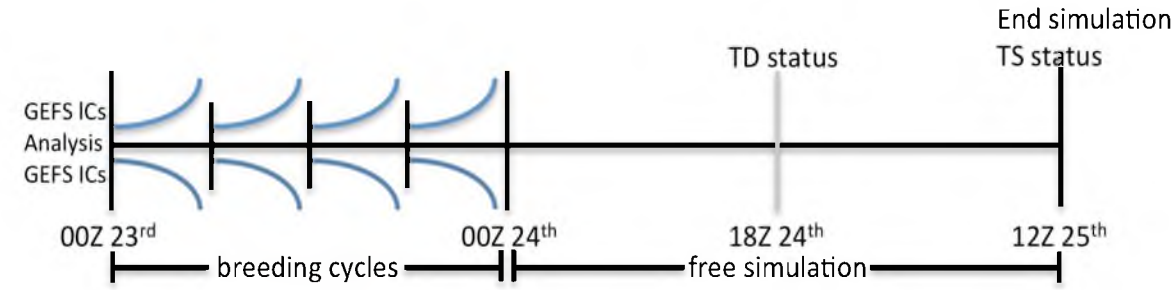

Figure 3. Diagram of the breeding cycles, which are started with GEFS initial perturbations, as well as the post-breeding simulation. This diagram shows 6 h cycling periods.

Table 2

Configurations of all initial condition-based ensemble simulations

Number	Ensemble name	Variable perturbed	BCs	Cycling period (h)	Localization
00	CNTL	Based on GEFS	FNL	N/A	N/A
01	FNLq3	Q	FNL	3	N
02	FNLth3	T	FNL	3	N
03	GEFSth3	T	FNL	3	N
04	FNLuvqt3	Q,T,U,V	FNL	3	N
05	FNLuvqt6	Q,T,U,V	FNL	6	N
06	FNLuvqt12	Q,T,U,V	FNL	12	N
07	FNLuvqt3_std	Q,T,U,V	FNL	3	N
08	FNLq3_height	Q	FNL	3	Y
09	FNLuv3	U,V	FNL	3	N
10	GEFSuvqt3	Q,T,U,V	FNL	3	N
11	FNLuvqt3_half	Q,T,U,V	FNL	3	N
12	FNLuvqt3_GEFSBCs	Q,T,U,V	GEFS	3	N
13	MULTI-PHYS	Physics	FNL	N/A	N/A
14	BS_PSIDOUBLE	T, stream	FNL	N/A	N/A
15	BS_VERT	T, stream	FNL	N/A	N/A
16	BS_CNTL	T, stream	FNL	N/A	N/A
17	BS_CNTL_NURI	T, stream	FNL	N/A	N/A
18	FNLq3_NURI	Q	FNL	N/A	N/A
19	FNLuvqt3_NURI	Q,T,U,V	FNL	N/A	N/A
20	BS_CNTL_WAVE3	Th, stream	FNL	N/A	N/A

not being perturbed but bred. Recall that the GEFS ICs have many fields that are perturbed.

The breeding period begins at 0000 UTC 23rd August and the breeding cycles are performed every 3, 6, and 12 h, with the cycling ending at 0000 UTC 24th. Thus, there are two to eight full cycling periods before we obtain initial perturbations for the ensemble forecast. In order to examine the impact of the perturbations from each variable, we choose to perturb various combinations of water vapor, potential temperature (θ hereafter), u, and v, and, when combined with the various cycling period possible, form the basis of the our breeding-related error growth research. Water vapor will here be defined as specific humidity (g/kg). Also, this perturbation scheme is a simple BGM method, as constructed by McLay et al. (2008), wherein a *spatially invariant*, or averaged, scaling factor is used to create perturbations from forecast errors at the end of each cycle period. Also, unlike many lower-resolution BGM configurations (McLay et al. 2008; Wei et al., 2008), here we only *add* the scaled perturbations (which are both positive and negative) to the analysis during the breeding, and do not double the number of members by subtracting them as well. This is due to the already-high computational cost of creating many ensembles of 14-member forecasts at high resolution (relative to our computational resources available), and is justified considering that when both adding and subtracting perturbations, one does not increase the effective degrees of freedom (EDF) of the ensemble (Wei et al. 2008).

Following Cheung and Elsberry (2002) and Snyder (2010), we define the genesis time as occurring when the NHC designates Ernesto a tropical depression, which was 1800 UTC 24th August 2006. The ensembles are begun 1 day before genesis, namely,

at 0000 UTC 24th August (since the GFS at that time only started every 12 h). After the 24 h breeding period, each ensemble forecast runs for 36 h, from 0000 UTC 24th to 1200 UTC 25th Aug 2006.

Statistical metrics and methods

The large output from the ensemble forecasts requires us to rely on several statistical methods in order to properly compare the various ensemble techniques and accompanying forecasts. The most basic of these is the ensemble mean, which is simply a mean across each of the 14 ensemble members in terms of a particular variable, or derived statistic. Relatedly, the ensemble spread is also calculated as the standard deviation of the ensemble members, which is calculated for either direct model variables or derived variables. The term bias, when used in this dissertation, represents the ensemble mean, in terms of variable being discussed, minus the corresponding analysis field, expressed as

$$\frac{1}{n}\sum_{k=1}^{n}(\overline{f}_k - a_k) \qquad (1)$$

where $\overline{f}$ denotes the ensemble mean of forecasts across the 14 ensemble members, *a* denotes the analysis, and *k* denotes the grid points across the domain.

In order to evaluate how the our ensemble forecasts compare in terms of the dimensionality of their error subspaces, we use the E-dimension statistic (E-dim). Patil et al. (2001) first introduced this metric to evaluate the NCEP GFS ensemble in terms of its important directions of variance. Originally called the bred vector dimension, this metric is based around empirical orthogonal functions (EOFs), which effectively reduce the

complicated nature of the matrix of interest into what are called principal components. This matrix, as explained below, is composed of ensemble output, in terms of the error of each member, in a particular configuration. Oczkowski et al. (2005) used the E-dimension to study low-dimensional development in certain regions of NCEP GFS model output, and explained that this E-dimension can describe the effective number of important directions in the "vector space spanned by ensemble perturbations by giving more weight to directions that explain larger portions of the total ensemble variance." Patil et al. (2001), expressed it as

$$\psi(\sigma_1, \sigma_2, \ldots, \sigma_k) = \frac{(\sum_{i=1}^{k} \sigma_i)^2}{\sum_{i=1}^{k} \sigma_i^2} \qquad (2)$$

where sigma represents the singular values of the matrix in question. Basically, the more unevenly distributed the variance, the lower the E-dimension.

In our analysis, we construct the matrix on which the EOF analysis is based by splitting the domain into square regions which contain an area of 10 x 10 grid points. This matrix is subsequently turned into a 100x1 vector, which we will call an E-vector, by putting each column in order below the first. Each of the 14 ensemble members contributes such a 100x1 vector, and our 100x14 matrix, *B*, is formed. This matrix is composed of data from the variable of one's choosing, in terms of each member's error, or distance from the analysis. In this dissertation, E-dimension refers to calculations based on member error, unless otherwise noted. A statistical function in Python computes the EOFs from this *B* matrix and provides the principal component coefficients, scores, and the eigenvectors of the matrix. The square roots of the eigenvalues yield the singular values which are used in the E-dimension calculation (equation 1 above). This is

equivalent to performing a singular value decomposition of the 100x14 matrix.

The interpretation of the E-dimension metric requires a more nuanced discussion than does its calculation. In the paper introducing the E-dimension, then called the bred vector dimension, Patil et al. (2001) described this metric as allowing them to identify an "effective dimension" spanned by the members of the ensemble of interest. Adopting an example from their paper, if 2 out of the 5 singular values are 0, in the *B* matrix, then the subspace spanned by the 5 local E-vectors, which represent ensemble members, is 3. Oczkowski et al. (2005) defined that E-dimension as characterizing the "effective number of dominant directions in the vector space spanned by the ensemble perturbations." Further, they explained that in a typical ensemble forecast, as the ensemble size is increased, the E-dimension is also expected to increase, until the number of members becomes sufficiently large to capture "all dynamically active phase space directions" in the region of study. The same authors went on to state that many low-dimensional areas correspond to troughs, ridges, and waves in the atmospheric flow, and thus that the location of low E-dimension regions does not necessarily just occur because of small ensemble size.

If, in a hypothetical ensemble, the members' data were drawn from a climatology for that particular variable, the E-dimension would equal the number of ensemble members minus one. By contrast, in actual ensembles, dynamical instabilities lead to dominant directions of uncertainty and thus substantially lower E-dimension. Patil et al. (2001) had originally found that regions of low E-dimension corresponded with regions of large estimated forecast uncertainties. They hypothesized that it is easier to improve model forecasts in such regions because of the correspondingly simple structure of the potential analysis and forecast error patterns in those areas. For example, Ockowski et al. (2005) found that the ensemble transform Kalman filter (ETKF; Bishop et al. 2001),

which was used during the 2000 Winter Storm Reconnaissance (WSR00) program to select optional dropsonde locations, nearly always selected regions of low-dimensionality. Kuhl et al. (2007) attributed similar findings as motivating the development of the local ensemble Kalman filter (LEKF; Ott et al. 2004) data assimilation scheme. Specifically, Kuhl et al. (2007) averred that the efficiency of the LEKF scheme was inversely proportional to E-dimension because of the fact that a negative correlation often exists between E-dimension and explained variance (i.e., the portion of error captured by the ensemble). Indeed, this had been noted by Szunyogh et al. (2005), who stated that when the local model dynamics were relatively more complex (i.e., high E-dimension), the Kalman filter could capture a smaller portion of the background error.

In terms of E-dimension characteristics especially relevant to this study, there are a few that should be noted. First, it has been found that E-dimension is generally higher in the Tropics compared with the extra-Tropics (Ockowski et al. 2005). As dynamical instability associated with baroclinicity is more prevalent in the extra-Tropics, low E-dimension in those regions is expected, although this depends locally on the respective growth rates of the unstable baroclinic modes. Relatedly, the same study found that E-dimension decreases with increasing forecast time because of the cumulative effects of dynamical instabilities, but is often followed by an unexplained increase in dimensionality near the end of the simulation. Overall, one can say that fast error growth generally leads to low E-dimension; however, the opposite is not true, as low E-dimension can also occur when forecast errors are small (Kuhl et al. 2007).

Basic evaluation of ensemble performance

In this section, a description of ensemble bias will be given as seen in all of the initial perturbation ensembles of this study, even though experiments describing some of them will not be detailed until later. This is done to provide a quick check of how our perturbation generation method in general compares with our control simulations. Figure 4 shows ensemble bias at 925 hPa throughout the simulation period for both water vapor and θ. Bias here is calculated, as in equation 1, by subtracting the (domain averaged) analysis from the ensemble mean at each time. Notice that at this 925 hPa height, all ensembles have a moist bias that reaches 1 g/kg and lasts throughout the simulation period. At the same height, most ensembles are slightly cool, but oscillate around the analysis.

Notice the bifurcation between ensembles at the starting time in Figure 4a. Those ensembles which have bred vapor show initial vapor values roughly 0.5 g/kg higher than the ensembles which simply used the FNL analysis, without any modification, to provide the initial vapor field, which are all zero in Figure 4a. Note that these differences are due to the breeding cycle itself and not because of differences between mean values in the FNL analysis at the different times at which FNLq3 and CNTL began. While FNLq3 starts breeding with ICs from 0000 UTC 23rd, recall that the rescaled perturbations are repositioned onto the analysis during each cycle, such that the FNLq3 and CNTL vapor fields are both influenced by values in the 0000 UTC 24th analysis. The same phenomenon is noticed in Figure 4b as well, where the ensembles breeding θ show initial θ fields about 0.3 K lower than those fields simply based on the FNL. However, despite these initial differences, within a few hours, the bifurcation is completely eliminated and

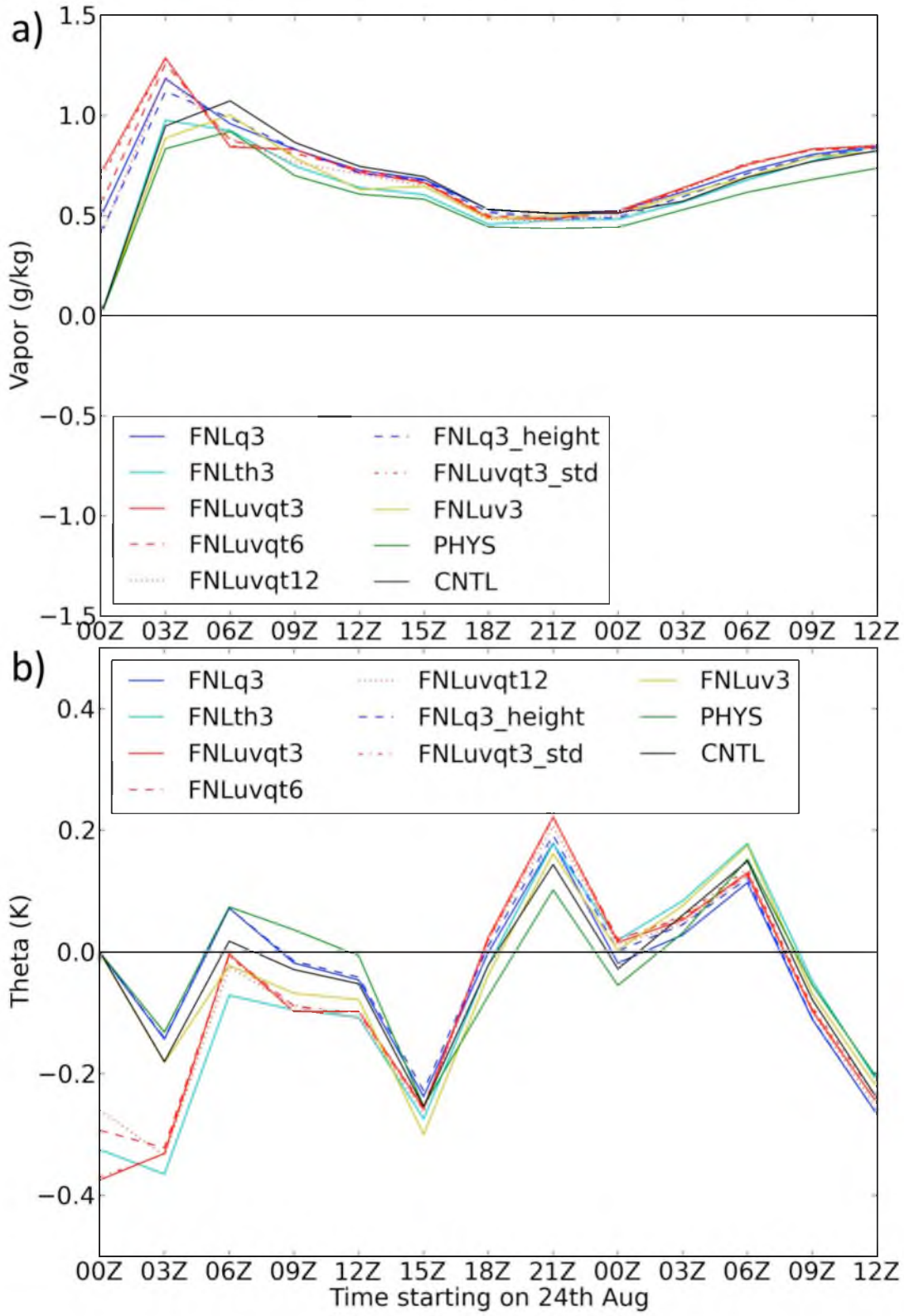

Figure 4. Time series of bias at 925 hPa for selected ensembles for a) vapor (g/kg) and b) theta (K).

the ensembles breeding vapor or θ then often show less bias than the control ensemble in terms of vapor or θ, respectively. This does not occur for every such case throughout the simulation period. However, it does happen often enough to show that initial differences from the FNL analysis, due to breeding, do not necessarily increase bias in that field going forward. In fact, compared to the control ensemble, the bred ensembles compare quite favorably.

Figure 5 provides the bias for all ensembles at 850 hPa. That the breeding of a particular variable has a disparate effect, depending on height, is shown if we compare Figure 5a with Figure 4a. At 925 hPa, breeding increased the amount of water vapor at 0000 UTC 24th compared to the FNL, whereas at 850 hPa, the breeding scheme produced notably less water vapor than the analysis. Note that at 850 hPa, it appears that those ensembles breeding vapor begin the free part of their simulations with 0.1-0.3 g/kg less water vapor compared with both the control ensemble and those not breeding water vapor. This difference with height is just as evident in terms of θ, as all ensembles breeding θ at 850 hPa start with much less of a cold bias compared with the corresponding values at 925 hPa. The 700 hPa ensemble bias values are provided in Figure 6, where we again see notably lower water vapor values in those ensembles which breed this variable compared to those ensembles that do not. At 700 hPa height (Figure 6), the breeding of both vapor and θ position those particular fields extremely well considering the drying and heating which occurs by 24 h into the simulation. The 500 hPa values are plotted in Figure 7 and demonstrate the same drying and warming which we see at 700 hPa. Overall, we see that inserting the bred fields into the FNL analysis at 0000 UTC 24th Aug creates ensembles with unique domain-averaged values, but that this

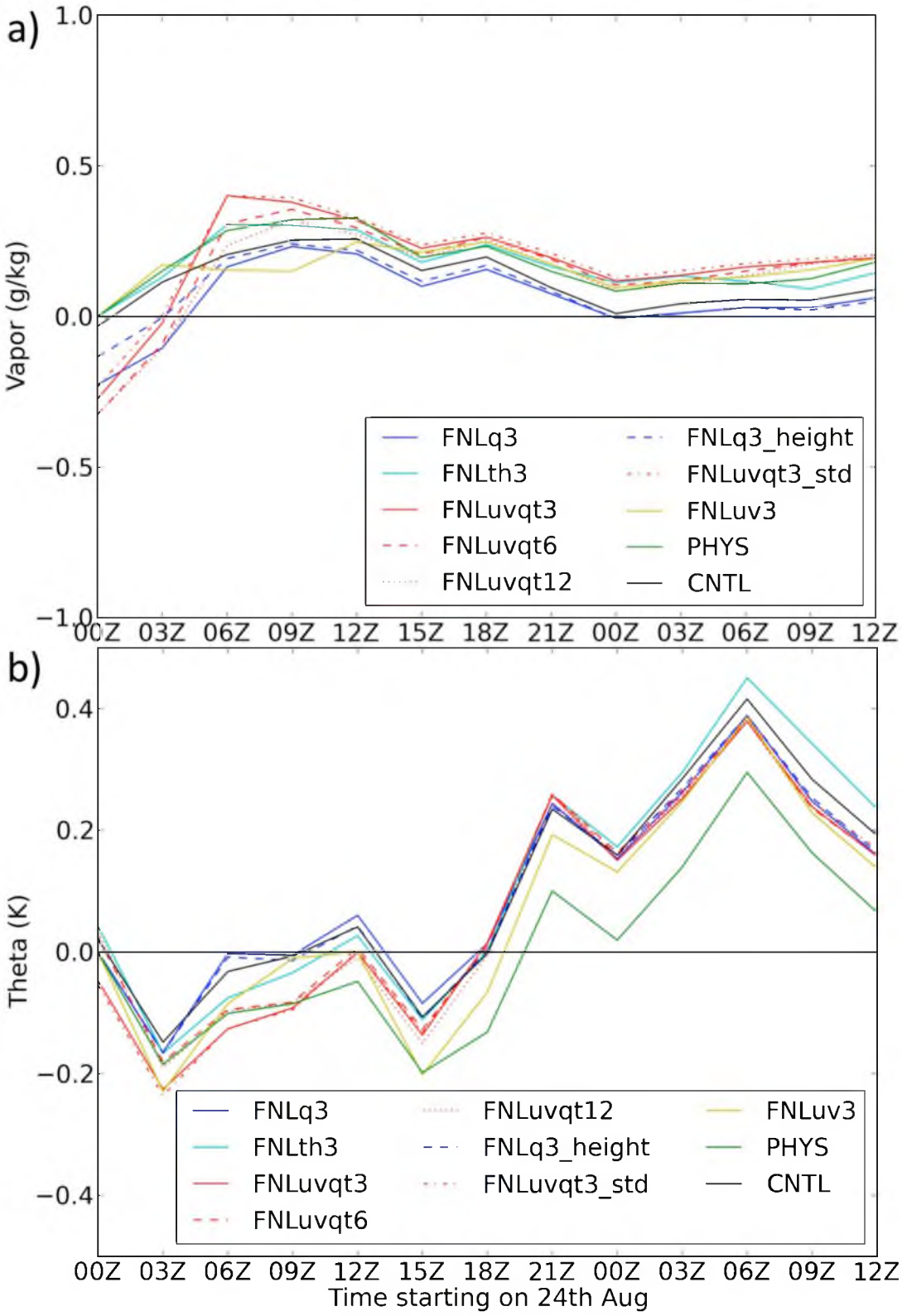

Figure 5. Time series of bias at 850 hPa for selected ensembles for a) vapor (g/kg) and b) theta (K).

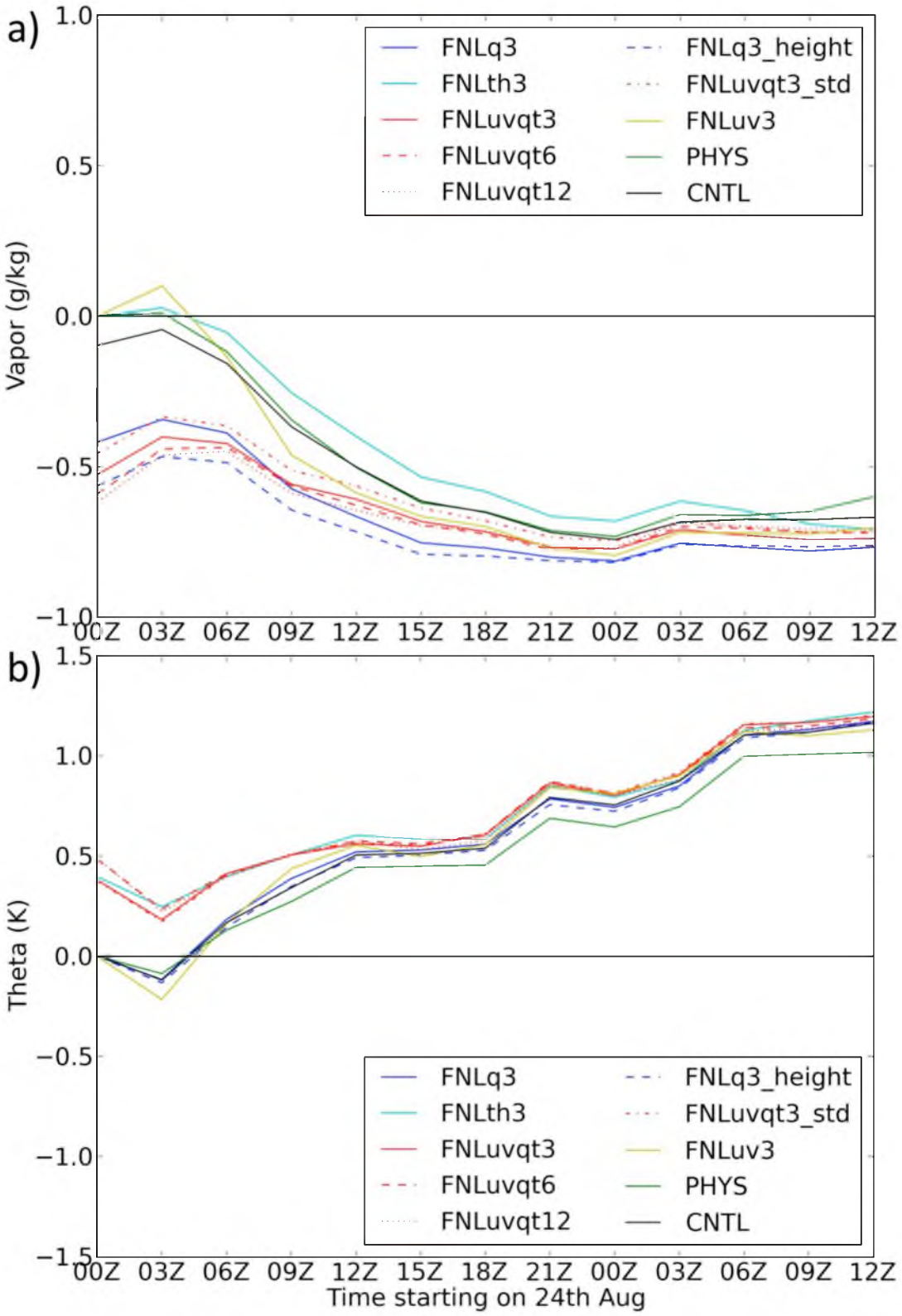

Figure 6. Time series of bias at 700 hPa for selected ensembles for a) vapor (g/kg) and b) theta (K).

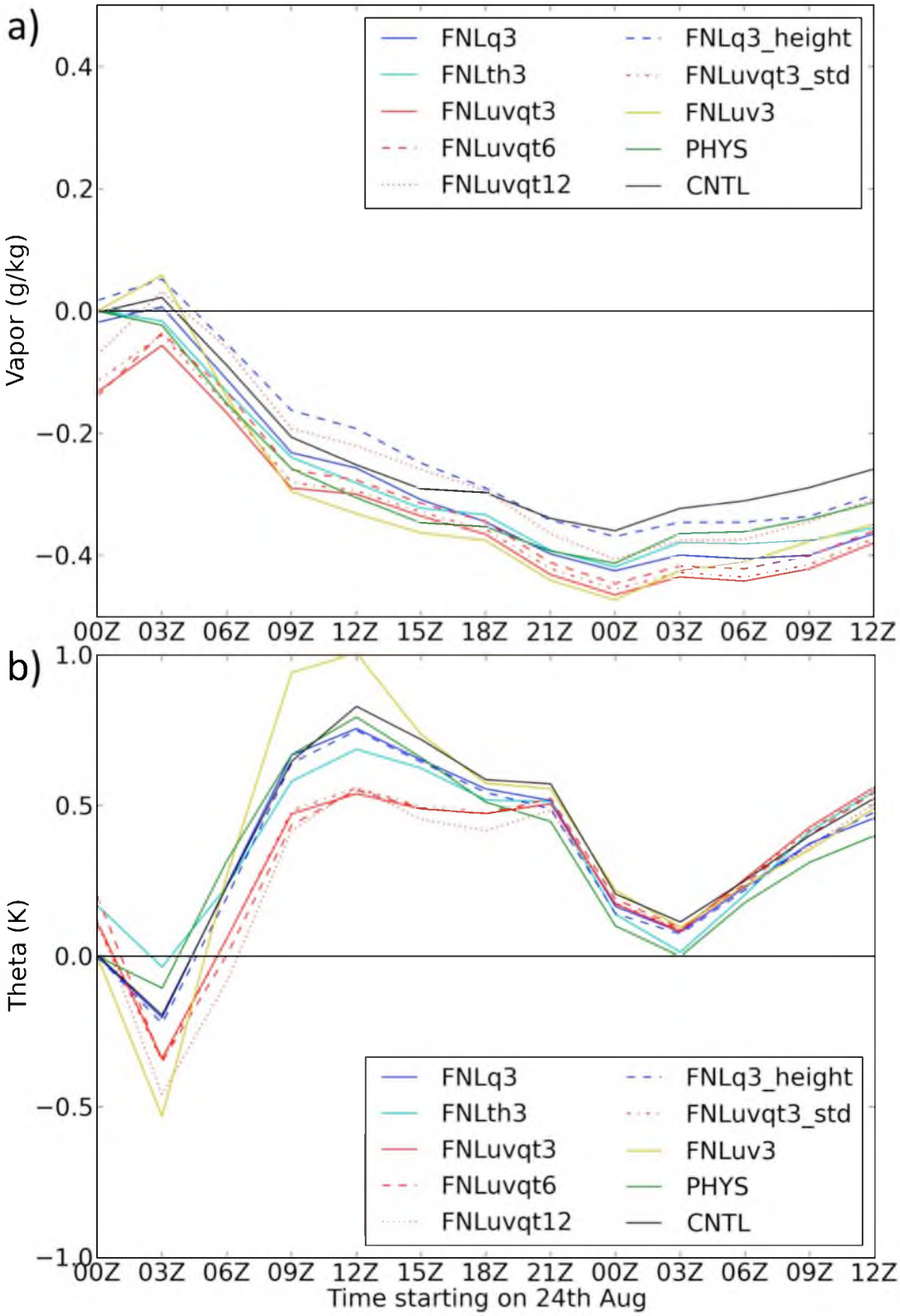

Figure 7. Time series of bias at 500 hPa for selected ensembles for a) vapor (g/kg) and b) theta (K).

does not lead to higher bias compared with the CNTL. In addition, the way in which vapor breeding produces notably lower averaged vapor values at 0000 UTC 24th compared to the FNL analysis, from 850 to 500 hPa, may have notable implications for the prediction of TC genesis, as water vapor at these levels has been noted for its role as a necessary genesis precursor (Bister and Emanuel 1997; Sippel and Zhang 2008). The effect of this will be examined later in the dissertation.

Ensemble bias by height from 1200 to 1800 UTC 24th is calculated and plotted in Figure 8. Within 50 hPa of the surface, we see that all ensembles have similar vapor bias, likely due to the effect of the same set of parameterization schemes across all ensembles. Elsewhere, however, we see that the various ensemble designs result in notable differences in bias with height, primarily in terms of θ values. For example, note how at most heights, the FNLuvqt ensembles show bias values 0.5 K less than those of the FNLth3 ensemble in terms of θ. While some of the interensemble differences are only of a few tenths of a Kelvin, recall that these calculations are averaged over the entire domain and a 6 h period, and thus demonstrate the unique ensemble characteristics resulting from the simple changes in breed ensemble design. While the difference between ensembles is less in terms of vapor, we note that the lowest vapor bias at certain heights is found in the FNLth3 ensemble, which demonstrates the fact that breeding a particular variable can cause notable differences not only in that same field, but also in those that interact with it. In general, we see that while the bias of the breeding ensembles is comparable to that of CNTL, the subtle changes in breeding technique lead to ensembles with unique forecast characteristics.

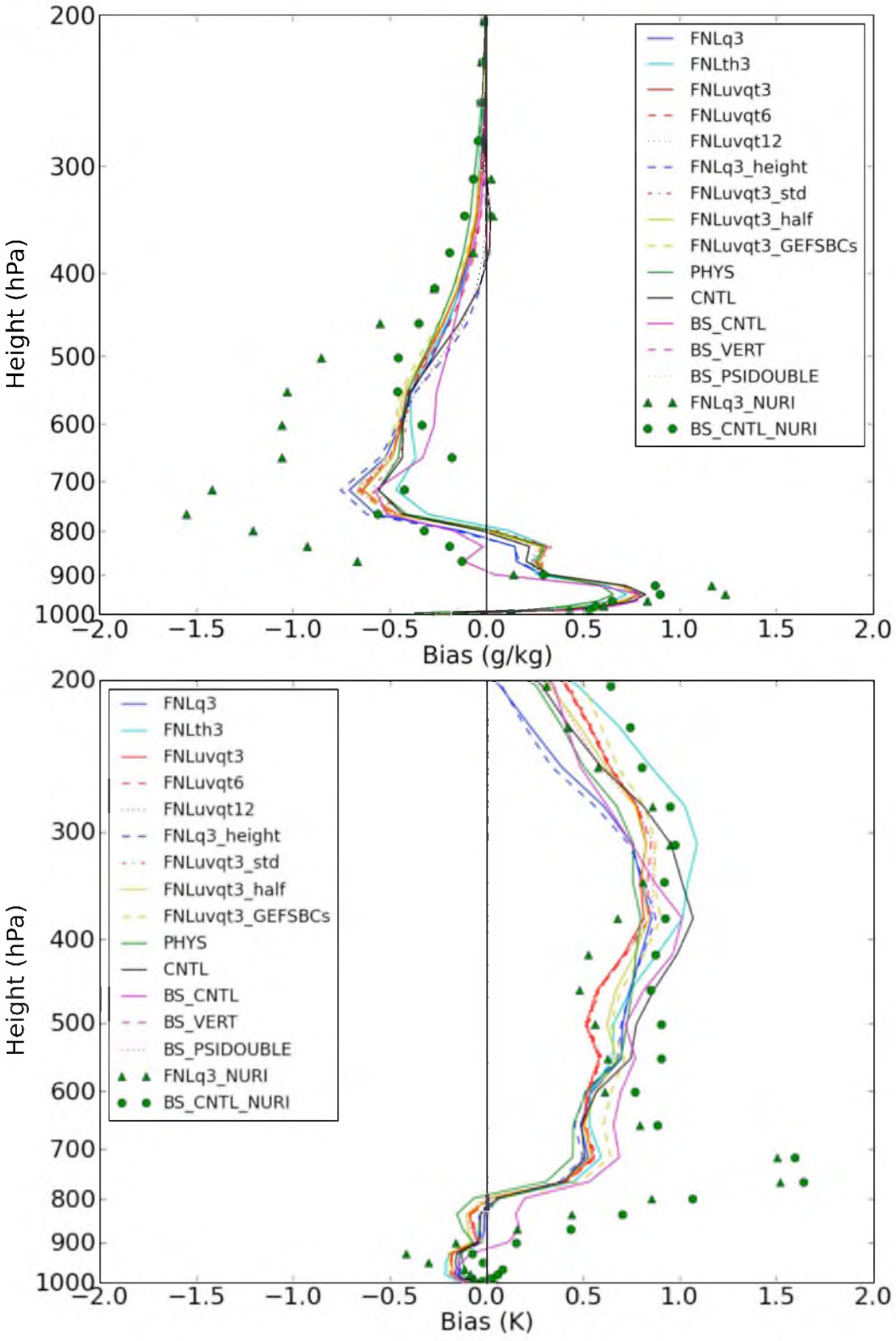

Figure 8. Bias by height from 1200-1800 UTC 24th August for selected ensembles for a) vapor (g/kg) and b) theta (K).

Regional GEFS-based ensemble (CNTL) versus global GEFS

We first evaluate the CNTL ensemble as it relates to the large-scale GEFS. Note that while the free part of the simulations, after the 24 h breeding cycle, began for all ensembles at 0000 UTC 24th August, the disturbance that becomes Hurricane Ernesto did not become a TD until 1800 UTC 24th. In addition, in this study, we define genesis as the transition to TD status, following Snyder et al. (2010). While this study does focus closely on ensemble error growth characteristics, the interaction between ensemble perturbation construction and genesis occurrence is also of great interest. This more physical focus, in addition to the statistical approach necessitated by our large amount of model output, is advantageous because of 1) the lack of particular studies on regional ensemble forecasts of TC genesis and 2) the fact that so much about this genesis process is still little understood.

With this in mind, and to highlight possible differences between global and regional ensemble forecasts, first we will describe the results of the GEFS forecast of Ernesto as it was created by NCEP. Fortunately, this was detailed by Snyder et al. (2010). Their study used GEFS data to study developing and nondeveloping TCs during NASA's African Monsoon Multidisciplinary Analyses (NAMMA) 2006 field campaign. Evaluating forecasts starting 1 to 3 days ahead of genesis, they determined the overall probabilities of genesis from the 14 GEFS members at each of these time steps. Snyder et al. (2010) found that the forecast 3 days ahead was for the pre-existing wave to remain weak, while the members were quite close together showing certainty that the wave would move (too) quickly to the west, "until the later stages of the forecast." To summarize their Ernesto-related intensity findings, at 1, 2, and 3 days lead time,

the GEFS predicted genesis in 4/14, 0/14, and 1/14 members, respectively. In general, forecasts started 1 day, as is done in our study, and 2 days ahead did produce more vortex circulations than did the forecast from 3 days before, but predictions of a well-developed TC still did not exist in any of the members (Snyder et al. 2010) and thus, the authors said that the ensemble forecasts of Ernesto largely "failed." Thus, this large-scale GEFS forecast of Ernesto will provide a difficult forecast baseline with which we will compare both our regional GEFS-based simulation, and the various implementations of our breeding scheme.

We now will examine synoptic views of the near-TC environment for the control run and several breeding ensembles, and then later examine the related differences in terms of genesis statistics. Plotted in Figure 9 are vorticity, wind vectors, and geopotential height values at 850 hPa for each of the members of the control ensemble at 1800 UTC 24th August. This time is chosen because it is when the proto-Ernesto reaches TD status. First, one quickly notices how intense the forecast of the disturbance already is in many members by this time. This contrasts with the fact that TDs are usually quite weak (minimum of 33 knot winds using the US 1-minute-average standard). Compared to what Snyder et al. (2010) found in the comparable GEFS results, it appears that this regional control ensemble (based on interpolated GEFS ICs) produces notably more intense storms, while also showing significant intensity spread among the ensemble members. For example, member 4, 5, and 7 predict TCs with minimum sea-level pressures (MSLP) 4 or 5 hPa (i.e., notably) below the 1009 hPa closed contour genesis threshold set in this study (and described below). Note that there is still a large degree of spread between the members in terms of TC intensity, positioning, and structure. For

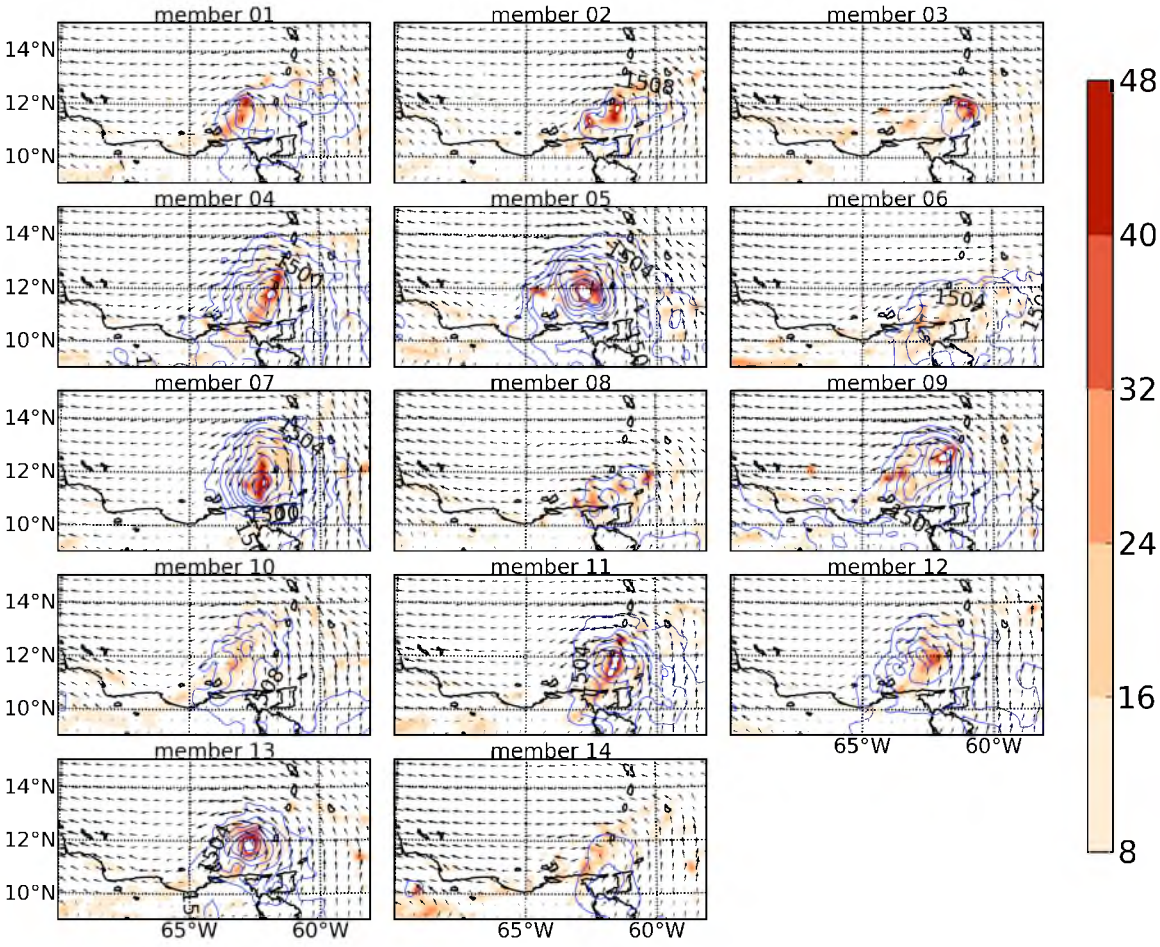

Figure 9. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of CNTL ensemble members at 850 hPa at 1800 UTC 24th.

example, member 3 shows a disturbance which does not even possess a closed wind circulation; compare this with the symmetric TC, reaching 1480 m in 850 hPa geopotential height, that is present in member 5 and 7. In addition to the TCs' high average intensity and intensity variability, also notice that the disturbances in member 12 and 14 are separated by more than 100 km after only 18 h, such that there is notable TC spatial variability among the members as well.

Thus, the CNTL, or *regional* GEFS ensemble not only produces forecasts of an intense Ernesto by best track genesis time, but also convey the large amount of uncertainty inherent in what Snyder et al. (2010) called a "difficult" TC forecast. While this forecast did provide much difficulty for the GEFS forecast on the large scale, these problems diminish substantially when rerunning it with a regional model.

CNTL versus breeding schemes

Having examined the genesis results of the CNTL ensemble, we now examine the relative differences between the CNTL and the bred ensembles with perturbations of one variable. Figure 10 presents a synoptic view of FNLq3 at the time of best track genesis. FNLq3 is the ensemble which is based around FNL initial conditions, where vapor is bred every *3* h and forms the sole perturbations of the 14 members. At first glance, we see that the majority of the simulated TCs in this ensemble are fairly well organized, with notable vorticity maxima and fairly symmetric circulation patterns. Note that, after this point, as TCs are mentioned in the context of our ensembles, they will be *simulated* TCs, and we will forgo the explicit mention of this each time. Most of the TCs in Figure 10 are less intense than those in the CNTL and in general appear to be just above the threshold of tropical depression status. The breeding and perturbation of vapor by itself appears

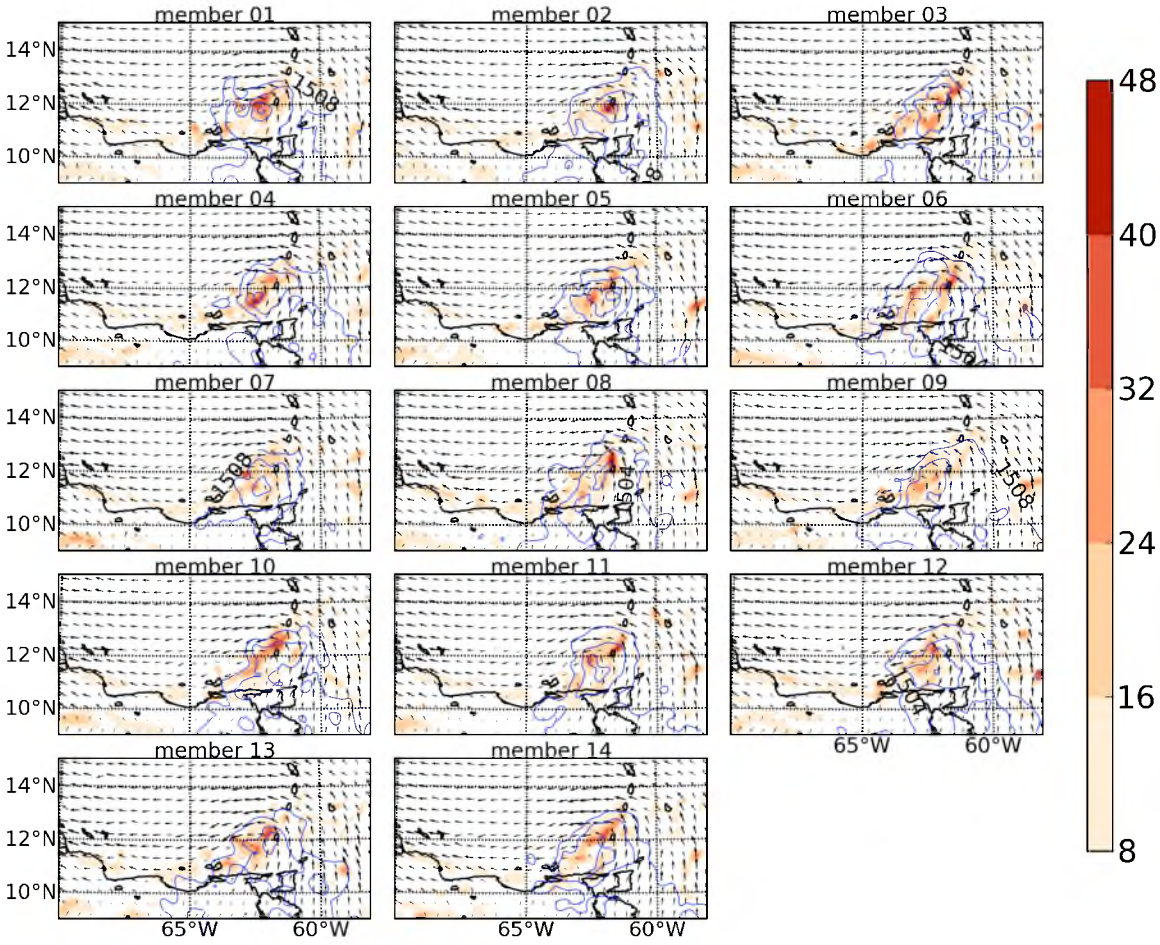

Figure 10. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLq3 ensemble members at 850 hPa at 1800 UTC 24th.

to have decreased the amount of TC intensity spread between the members of the ensemble compared with the control ensemble. Since the CNTL and FNLq3 IC vapor perturbations have similar intermember variability, the TC intensity spread differences at 1800 UTC demonstrate the impact between perturbing vapor (FNLq3) on one hand, and several variables (which occurs in CNTL) on the other.

Presented in Figure 11 is the synoptic view of the FNLth3 ensemble. We quickly notice that these TCs are notably more intense than those of the FNLq3 ensemble and many appear better organized than the corresponding members in the CNTL ensemble. In addition, we see that the intensity variability is fairly small in the FNLth3 ensemble and this result again demonstrates the impact of just perturbing θ (in FNLth3) compared with an ensemble with many perturbed variables (CNTL). That the members of CNTL and FNLth3 show notably stronger TCs than in FNLq3 may relate to the fact that average 7-800 hPa water vapor at the initial time in the latter ensemble is notably reduced compared to those that did not breed vapor (Figure 5 and 6). These physical precursors to genesis will be discussed in more depth later.

To summarize the genesis characteristics of these three regional ensembles (namely, CNTL, FNLq3, and FNLth3) and the global GEFS ensemble, statistics are presented in Table 3. In terms of the genesis criteria for regional ensemble forecasts, genesis is said to occur in a member if its forecast contains a MSLP contour of 1009 hPa or below and a closed 850 hPa geopotential contour of 1496 m or below, both of which must be found within, or overlap, a closed wind circulation at 850 hPa. To be clear, for genesis to be counted as occurring at 1800 UTC 24th, these characteristics must be present at that time. Genesis criteria from Snyder et al. (2010) are similar, but include a

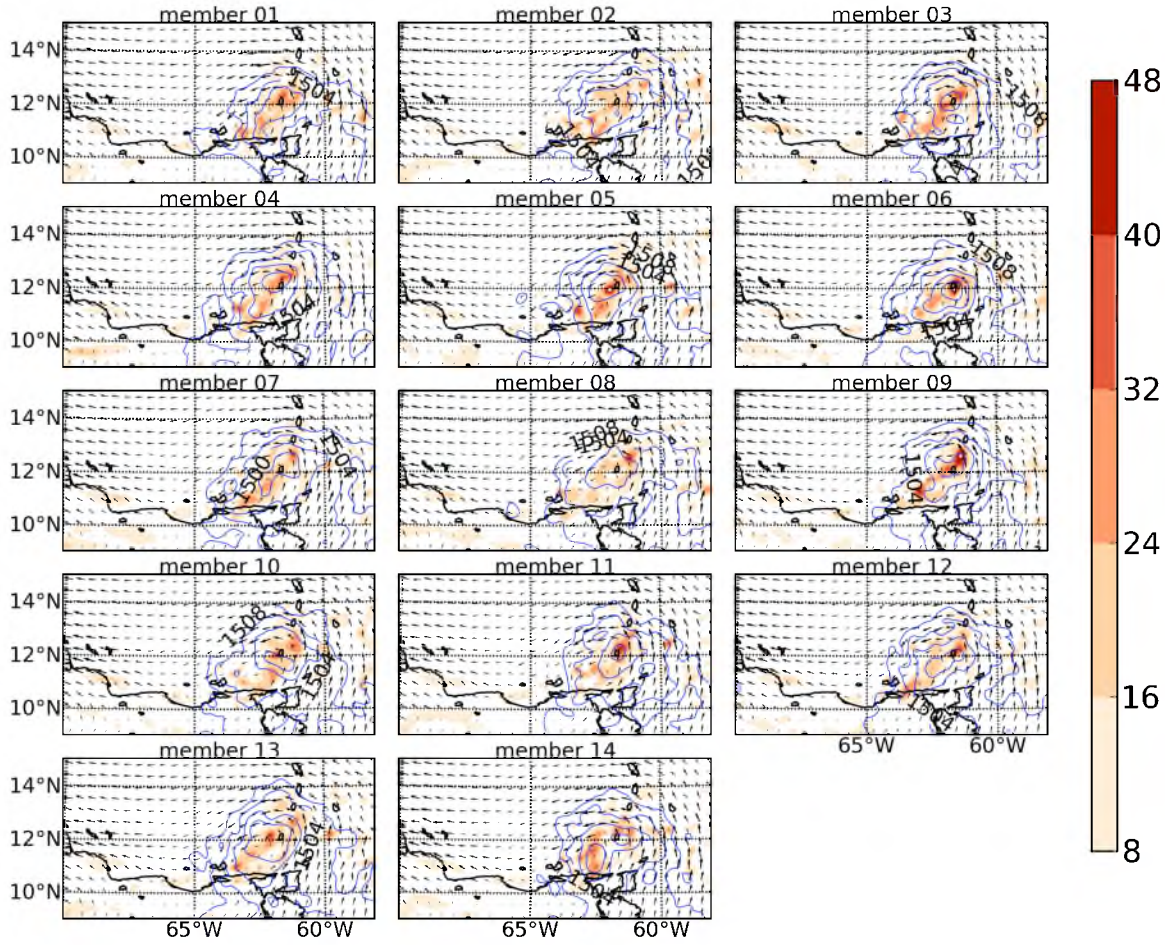

Figure 11. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLth3 ensemble members at 850 hPa at 1800 UTC 24th.

Table 3

Genesis statistics for

Hurricane Ernesto (forecast began 0000 UTC 24th)

Ensemble experiments	18Z 24th	12Z 25th
GEFS global	04/14	n/a
GEFS regional (CNTL)	07/14	13/14
FNLq3	02/14	14/14
FNLth3	12/14	14/14

warm core threshold and more contour specificity, which is necessary when using a lower resolution ensemble. Table 3 uses the regional criteria specified above and includes statistics for the GEFS as calculated by Snyder et al. (2010). Note that Snyder et al. (2010) did not produce global GEFS genesis results for 1200 UTC 25th Aug.

As expected from the preceding analysis, in Table 3, we note that the regional CNTL WRF-based ensemble, based on GEFS IC interpolation, produces almost twice as many genesis cases by 1800 UTC as does the GEFS global ensemble. As other studies have mentioned, it appears that genesis occurs more easily when the relevant processes are explicitly resolved (Hennon and Hobgood 2003), as is more likely the case in our regional ensemble. The other notable point of Table 3 is the fact that the FNLth3 ensemble produced many more genesis cases by the time of best track genesis than the CNTL or FNLq3 ensemble. Thus, in this preliminary check on breeding ensemble outcomes, we find that the results produce well-formed vortices similar in structure and strength to those of the basic CNTL ensemble. Interestingly, by 36 hours into the simulation, each of the bred the ensembles produces genesis in each of their members. This allows us to focus on the possible impact of genesis-related precursors from 0000 UTC to 1800 UTC 24th Aug, since this period produces notable differences in genesis formation. In the next chapter, we will look at the way initial conditions and error growth might influence these differences, and later in this paper, we will use statistical methods to analyze the possible impact of variables such as vapor, θ, vertical shear, and vorticity on genesis rates within this 0000 UTC to 1800 UTC 24th Aug time window.

Highlights of Chapter 2

- The WRF-based regional ensemble forecasts (CNTL), downscaled from the GEFS (global ensemble) ICs, better predict genesis compared with the global GEFS forecast.
- Depending on the details of the scheme, the regional breeding ensemble forecasts can produce higher probabilities of genesis than the regional ensemble downscaled from the GEFS ensemble (CNTL).

The results imply that regional mesoscale ensembles can notably enhance the prediction of TC genesis.

CHAPTER 3

INITIAL PERTURBATIONS, ERROR GROWTH, AND DIMENSIONALITY

Error growth: Potential temperature versus vapor

Owing to the obvious contrasts in physical implications, the ensembles with individual initial perturbations of vapor and θ (FNLq3 and FNLth3, as evaluated in the previous chapter) are the two experiments that we will compare in depth, although later we will look at more complicated breeding configurations.

In order to provide an example of how error growth evolves as a variable is bred over time, Table 4 contains the overall scaling factors for the FNLq3 and FNLth3 ensembles' first member throughout the 24 h cycling period. First, recall that this scaling factor is based on an average over the domain, and over all heights, for both water vapor and θ. As the scaling factor is multiplicative, and used to bring the bred error magnitude at the current time back to the magnitude of the errors 3, 6, or 12 h prior, a larger scaling factor means a smaller amount of error growth over the last 3 h period. Small factors thus denote faster error growth.

In terms of the differences between breeding vapor versus θ, quite consistently, θ scaling factors are smaller than those of water vapor, which is true throughout the simulation period and demonstrates the fact that in convective environments θ errors grow notably faster than vapor error. The time when the error growth rates plateau for the

Table 4

Comparison of scaling factors for member 1 of

FNLq3, FNLth3, FNLuv3 ensembles during 24 h breeding period

Time	FNLq3	FNLth3	FNLuv3 U	FNLuv3 V
03Z	0.83	0.80	0.96	0.97
06Z	0.75	0.67	0.83	0.87
09Z	0.77	0.64	0.84	0.80
12Z	0.76	0.54	0.75	0.68
15Z	0.73	0.58	0.80	0.68
18Z	0.66	0.58	0.74	0.66
21Z	0.73	0.59	0.73	0.67
00Z	0.77	0.62	0.71	0.65

two variables is notably different as well, as θ error growth rates stop increasing around 12 h into the breeding period, whereas vapor error growth rates are still increasing (albeit only slowly) 18 h into the period. What we find from these scaling facts is that, at their peak, vapor error grows by about 50% (1 - 1/.66) over a 3 h period, whereas θ error grows by roughly 85% in 3 h (1 – 1/.54). Later, we will discuss the impact of the specific cycling period and compare results with both 6 and 12 h cycles. Next, we will examine the results of the faster θ versus vapor error growth in terms of evolving error structures and final perturbations.

In order to view the regional differences masked by the domain averaging done in Table 4, Figure 12 shows the 850 hPa bias, again for the first member of the FNLq3 and FNLth3 ensembles, every 3 h throughout the 24 h breeding period. Here we see very dynamic error growth in terms of both θ (Figure 12a) and vapor (Figure 12b), in that there are notable error structures in each that arise for a few hours and then disappear. The breeding cycles thus do not just simply amplify or reduce already existing perturbations provided by the GEFS, but, according to the flow, amplify the fastest growing modes and provide unique regional anomalies quite different from those present at the beginning. In order to demonstrate the differences between scaled and nonscaled fields, the last time step exhibits the field right after the error rescaling, while the fields above it show the unscaled fields, or, the fields 3 h of error growth. During the whole 24 h period local areas of vapor bias remain within 2.5 g/kg and local areas of θ stay within roughly 2 K of the analysis. The scaled fields appear to be slightly above half of those values, which aligns with the general pattern found in Table 4.

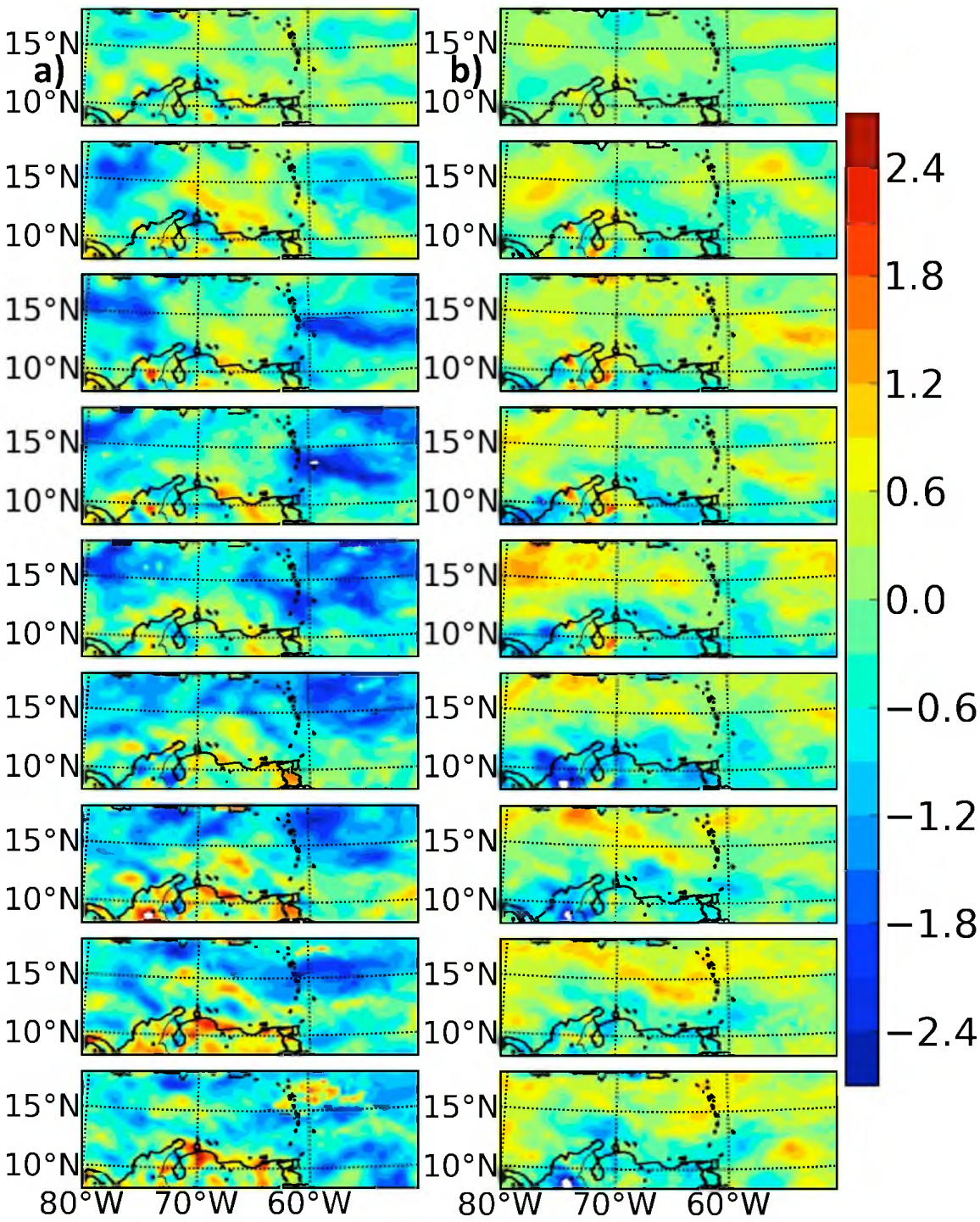

Figure 12. Member 1 bias at 850 hPa for a) FNLq3 ensemble in terms of vapor (g/kg) and b) FNLth3 ensemble in terms of theta (K). Bias is calculated every 3 h from 0000 UTC 23rd to 0000 UTC 24th. The same colorbar applies to a) and b). The last time step (bottom panels) display fields after the rescaling, while the others display fields *before* the rescaling.

Differences in initial perturbations: CNTL, FNLq3, and FNLth3

In order to provide context for the 14-member ensemble initial conditions after θ and vapor breeding, we first look at the characteristics of θ and vapor in the control ensemble, which starts simply with GEFS ICs at 0000 UTC 24th. Plotted in Figure 13 are initial 850 hPa vapor bias values for each of the 14 members in the ensemble. Note that in most regions, the perturbations are confined within 1 g/kg of the analysis. While, at first glance, there appear to be notable small-scale differences among the members, upon inspection, each of the 14 members generally show enhanced water vapor levels on the far western side of the domain, and also in the middle of the domain around 13°N and 60°W. Similarly, Figure 14 shows the control ensemble's θ perturbations from the analysis at 0000 UTC 24th at 850 hPa. Compared with water vapor, the control ensemble shows notably higher spatial variability among its members in terms of θ. Note that θ perturbations only rarely exceed 1 K, and this occurs in different locations for each member. While one has to be careful when comparing two different physical quantities like this, it can be said that in one field (θ; Figure 14), the domain maxima vary in terms of spatial positioning among the members to a higher degree than it does in another field (vapor; Figure 13).

Recall that the domain-averaged 850 hPa vapor for the FNLq3 ensemble ends the breeding period (at 0000 UTC 24th) with a notable dry bias (Figure 5a). In order to view spatial differences and compare these with the control ensemble, in Figure 15 FNLq3 850 hPa water vapor bias at 0000 UTC 24th is plotted for each of the 14 ensemble members. Indeed, in this ensemble, much of the domain denotes a dry bias compared with the analysis, which helps confirm the accuracy of the various calculations. Not only this,

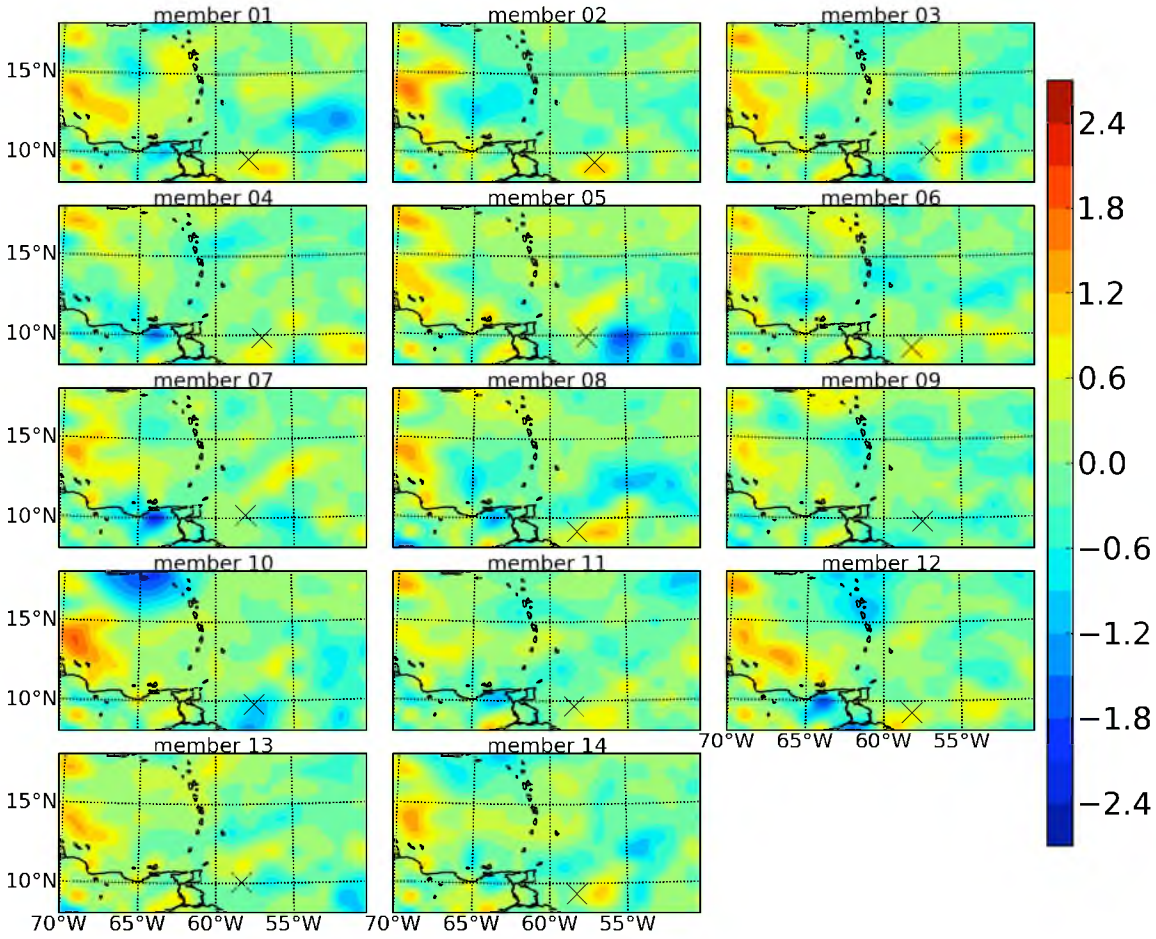

Figure 13. Bias of 850 hPa vapor (g/kg) for each member of CNTL ensemble at 0000 UTC 24th. X marks the location of the pre-Ernesto disturbance in each member.

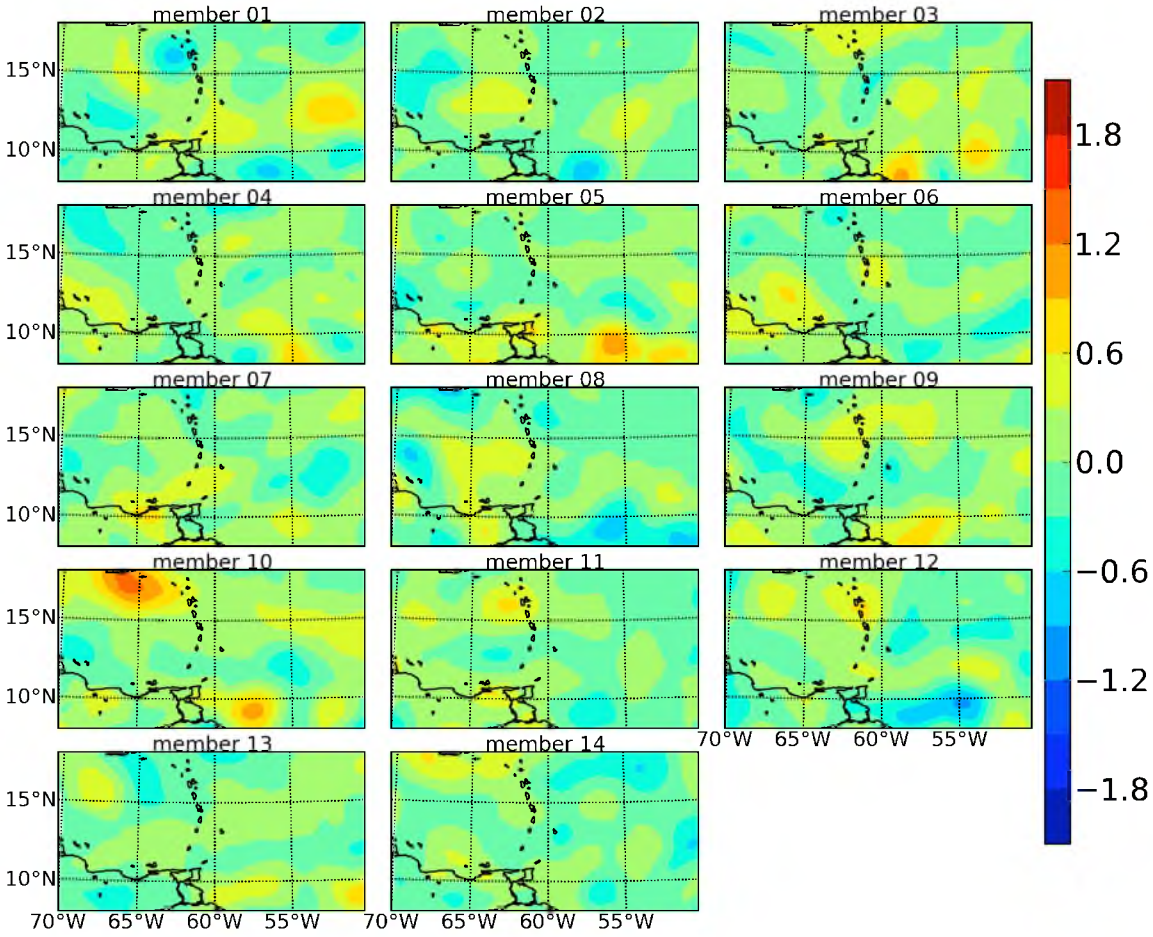

Figure 14. Bias of 850 hPa theta (K) for each member of CNTL ensemble at 0000 UTC 24th.

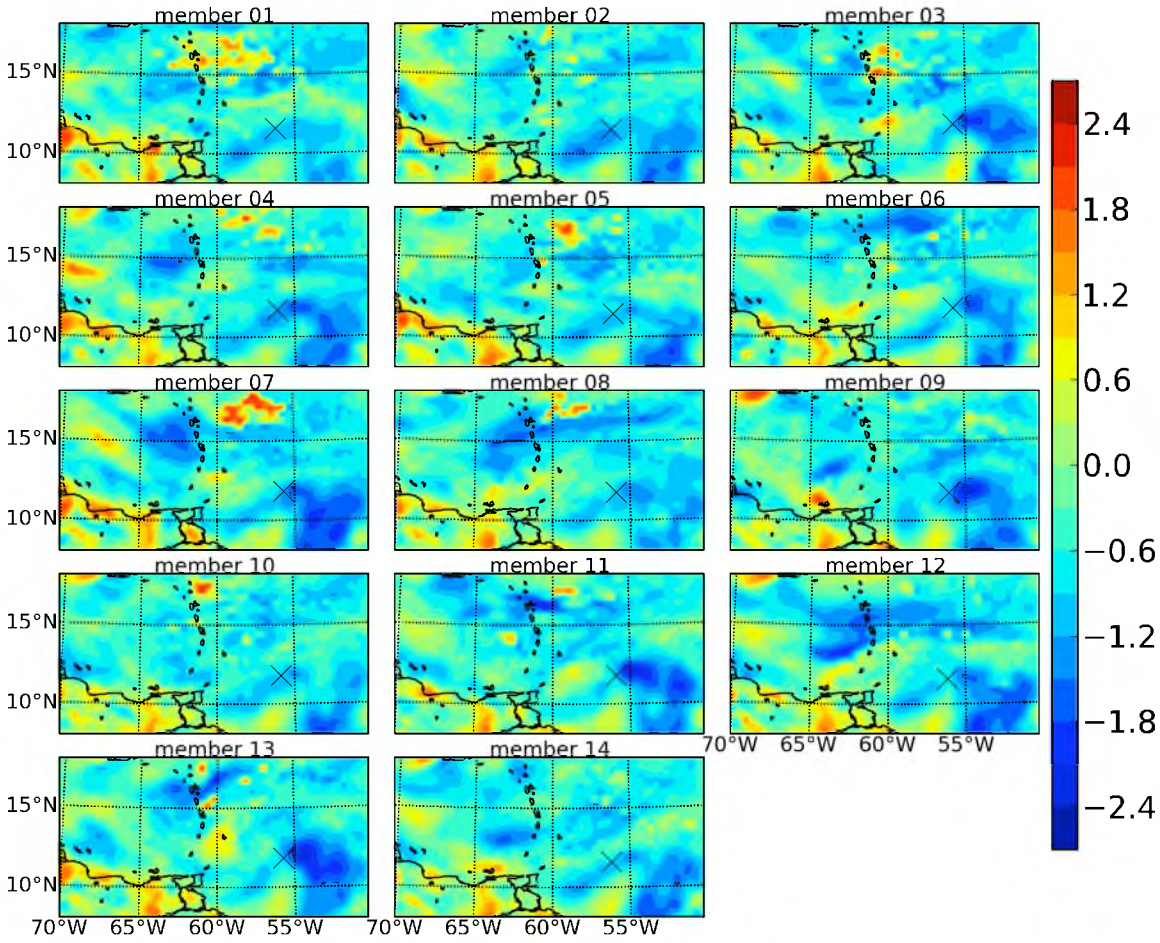

Figure 15. Bias of 850 hPa vapor (g/kg) for each member of FNLq3 ensemble at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

but these dry biases are in very similar positions among the members, which regional *intra-ensemble similarity* is akin to what we see in the control ensemble's vapor distributions (Figure 13). Note how the dry areas in the FNLq3 ensemble (Figure 15) are mostly concentrated in the southeastern quarter of the domain. Comparing Figure 15 and 13, it is striking to compare the much greater dry bias of the FNLq3 ensemble with that of the control ensemble. Note that these differences are just due to the BGM and ET (which CNTL is based on) schemes themselves. While the breeding ensemble does begin with ICs from 0000 UTC 23rd, during the cycling process, the BGM method eventually updates values according to what is present at 0000 UTC 24th.

In terms of how these IC differences cause genesis differences between FNLq3 and FNLth3, it is important to recall that in FNLth3, the vapor field is simply taken from the FNL analysis on which the CNTL is based. Thus, the values of FNLq3 bias (Figure 15) show how vapor actually varies between FNLq3 and FNLth3. Since the bias is here calculated as the vapor in FNLq3 minus that in the analysis (or FNLq3 vapor minus FNLth3 vapor), it is easy to see that the FNLq3 ensemble has large areas of reduced water vapor in the critical 850 hPa layer compared with FNLth3. This difference between FNLq3 and FNLth3 is most notable in the southeast corner of the domain for each member, and likely affects the local pre-TC environment, *which is in this southeast corner*, such that genesis is inhibited. A clear difference in initial domain-averaged water vapor between FNLth3 and FNLq3 at 850 hPa is also seen in Figure 5a. Recall that the FNLq3 ensemble shows only two members with genesis, by the actual genesis time, compared to 12 members in the FNLth3 ensemble, which appears to demonstrate a notable water vapor impact on the early organizational phase. The sizeable impact of this

water vapor on genesis differs from that usually described in the literature (McBride and Zehr 1981), where the Tropics at certain times of year are described as having the water vapor necessary for genesis, but usually lack the pre-existing vorticity and/or low vertical shear levels in order for genesis to occur. Here we see almost the opposite, wherein pre-existing wave is present (in the FNL downscaled analysis; not shown). However, local 1-2 g/kg vapor differences appear to notably alter the likelihood of genesis.

Figure 16 shows θ bias values in the FNLth3 ensemble at 0000 UTC 24th Aug, or, just after the last error rescaling associated with the breeding period. Across the domain, areas of warm bias slightly outnumber those showing a cold bias, with perturbations remaining largely within 1.5 K of the analysis. Notice the similarities between perturbations in θ and vapor (Figure 15), in that the high θ areas often collocate with areas of low moisture. This pattern holds quite consistently across the 14 members. Comparing these bred θ members to those in the control ensemble (Figure 14), we see that the perturbations associated with FNLth3 (Figure 16) have notably higher magnitudes than those of the CNTL, with the domain-averaged bias of the two ensembles being very similar at this time (Figure 5b). Amongst the FNLth3 members (Figure 16), θ perturbations show more significant intra-ensemble spatial differences compared to the FNLq3 ensemble's vapor field (Figure 15), which is similar to what we find in the corresponding spatial patterns of the control ensemble at this time. From this brief analysis of IC fields at 0000 UTC 24th, it appears that vapor spatial structures among the ensemble members develop fairly similarly, whereas those of θ show more notable (inter-member) differences across the domain. Interestingly, this pattern holds for both bred fields and those derived straight from the GEFS and its ET scheme, such that it appears to

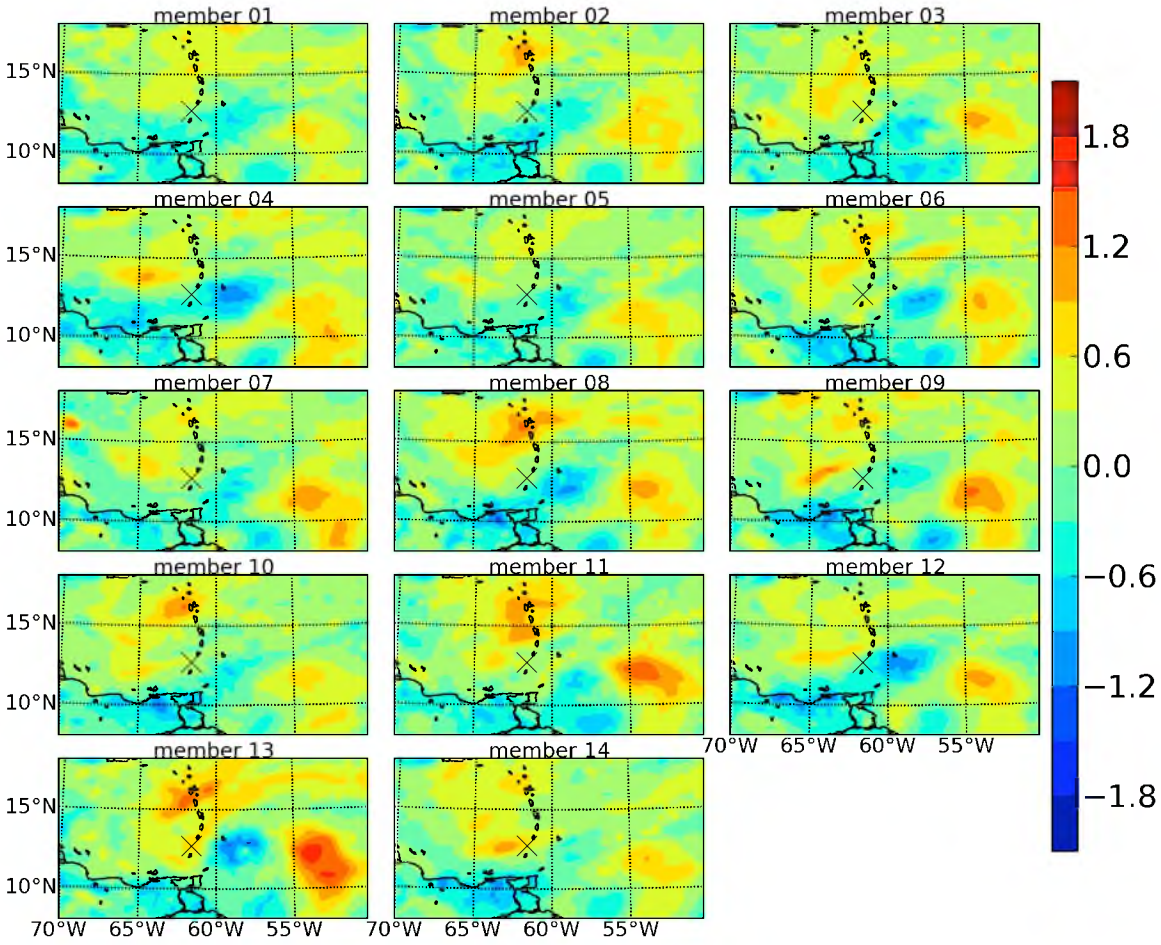

Figure 16. Bias of 850 hPa theta (K) for each member of FNLth3 ensemble at 0000 UTC 24th (end of breeding period).

describe characteristic differences between θ and vapor in the Tropics. Finally, we find that differences in initial water vapor levels appears to have a significant impact on the likelihood of genesis between the FNLq3 and FNLth3 ensembles.

E-dimension and ensemble spread

Having discussed the genesis and error growth differences in the breeding schemes compared with CNTL, we now turn to statistical methods in order to understand the more nuanced differences between the ensembles. Plotted in Figure 17 is domain-averaged 850 hPa water vapor error E-dimension and spread over time for the CNTL, FNLq3, and FNLth3 ensembles. Essentially all operational ensembles are underdispersive (Berner et al. 2009), such that they do not account for all the possible uncertainty implicit in their forecast. Both in operational and research settings, the use of multiple models is a common way of mitigating this underdispersion problem (Hagedorn et al. 2005), and this phenomenon is often measured by comparing ensemble spread and error. Considering our E-dimension data, we think it fitting to similarly compare an ensemble's forecast uncertainty with its forecast error dimensionality. In Figure 17, spread is calculated as the standard deviation over the ensemble members. While the axis limits in this figure could be tightened, they are set such that easy comparisons can be made with later figures. In terms of E-dimension, the first thing to notice is that while the values for the three ensembles are quite similar over much of the simulation period, the bred ensembles achieve higher E-dimension values more quickly, up to 0900 UTC, compared with the control ensemble. Between the vapor and θ breeding, however, there are few if any differences before 12 h. After this time, the θ breeding ensemble produces consistently, but just slightly lower E-dimension than both the FNLq3 and control ensembles.

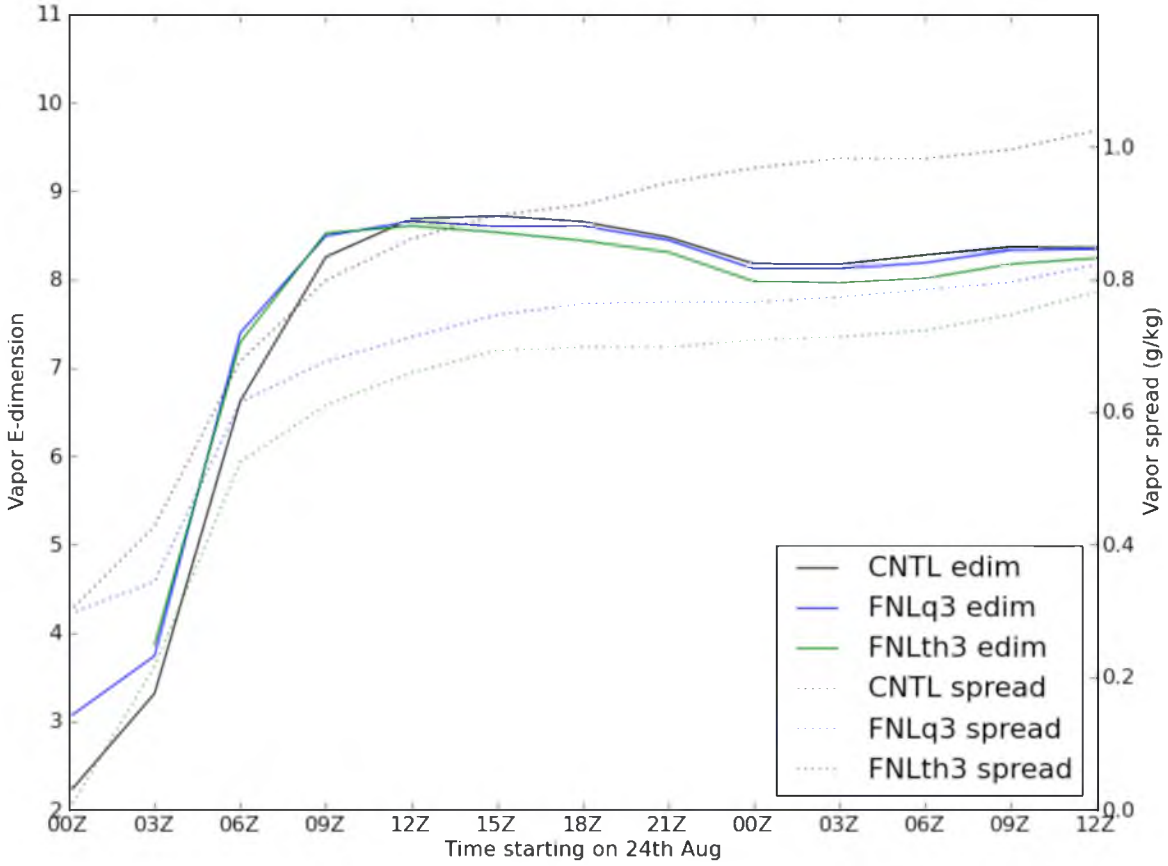

Figure 17. Time series of error E-dimension and spread (g/kg) for 850 hPa domain-averaged water vapor in the CNTL, FNLq3, and FNLth3 ensembles.

Interestingly, there is only a very slight reduction towards low E-dimension, from 24-36 h, compared to that described by Oczkowski et al. (2005); compare with their Figure 2. This is likely due to the fact that here we use a regional model in the Tropics, whereas they were using the global GFS in the extra-Tropics where errors, and thus low-levels of E-dimension, related to baroclinic energy conversion are more common.

Note that the two bred ensembles show notably lower vapor *spread* after 6 h than does the control ensemble (Figure 17). While this is to be expected in the FNLth3 ensemble because the only initial variance is in terms of θ, this reduced spread is present in the FNLq3 ensemble also. The latter case may be explained in terms of the fact that here, only one variable is perturbed at the start time compared with many variables in the control ensemble. Over time, these other perturbed variables allow for more variability in the vapor fields in the CNTL compared with FNLq3 ensemble. While reduced variance is expected when only perturbing one variable versus several, these results quantify the extent to which this occurs in high-resolution simulations.

In order to view localized differences between ensembles in terms of the establishment of dimensionality, a plan view of E-dimension is shown in Figure 18 over our geographic area of interest and for the three ensembles of interest just 6 h into the simulation. In addition to water vapor error E-dimension at 850 hPa, that of θ and wind is also shown at the same height. Wind E-dimension is here calculated by first finding the magnitude of the wind at each grid point and then calculating the error dimensionality of those values minus that of the analysis. First, notice how the E-dimension is notably higher over much of the domain in the bred ensembles (Figure 18b,c) compared with the control ensemble (Figure 18a). This feature is present for water vapor, θ, and wind. This means that the dimensionality of these variables grows notably

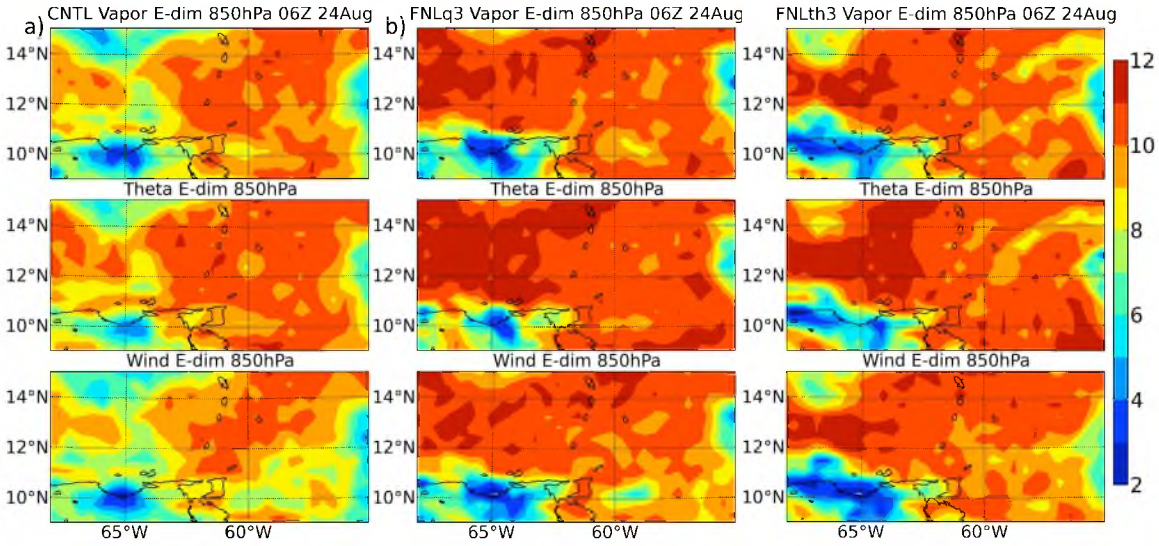

Figure 18. Error E-dimension for water vapor, theta, and wind at 850 hPa at 0600 UTC 24th August for the a) CNTL, b) FNLq3, and c) FNLth3 ensembles.

more quickly in the first 6 h of the simulation in the bred ensembles compared with the CNTL ensemble. Essentially, by 0600 UTC 24th, the bred ensembles show more variance in more error directions compared with the control ensemble. Note that this is true as well in Figure 17 at 0600 UTC 24th, although the differences appear smaller there due to the domain averaging. While this pattern holds over most of the domain in Figure 18, it is especially notable over the ocean. Near the Venezuelan coast, however, the breeding methods show a notably lower E-dimension for vapor, θ, and wind compared with the CNTL ensemble. This is possibly due to complications surrounding the interaction of the bred error fields with land-surface and radiation parameterization schemes after the bred fields are placed in the FNL analysis at 0000 UTC 24th August. As the behavior and error in each of the members would likely be more tied to these parameterizations in the Venezuelan region, which all members have in common, compared with the open ocean, the directions of error variance over that area substantially decrease.

While one would think that an ensemble should have as many error variance directions as ensemble members (McLay et al. 2008), in reality, error variance directions are not so simple to interpret. Any error dimensionality discussion requires a full disclosure of the related caveats. Here it might be helpful to step back from these E-dimension comparisons and discuss whether a high or low number of error variance directions is something desirable in ensemble evaluation. First off, ensemble members are created to help span the subspace of possible analysis error, or to describe the corresponding probability density function in multidimensional phase space (Epstein 1969). As explained, higher E-dimension means more variance spread across more error directions, compared to low E-dimension. We must note that Wang and Bishop (2003), Wei et al. (2006), and McLay et al. (2008) criticized certain breeding implementations for

producing analysis perturbations whose variance was concentrated in notably fewer eigendirections than the number of perturbations. Despite this, however, one cannot simply interpret a smaller difference between E-dimension and the corresponding number of ensemble members as being an indicator of a "good" ensemble. To render such judgments, one would first have to determine whether the ensemble is capturing the space in which the forecast error evolves or whether the forecast error is outside of the ensemble space (Kuhl et al. 2007). Because of this, when discussing the E-dimension results, we will refrain from saying that high values are "better" than low values and will instead discuss the differences in error dimensionality between various ensemble creation methods. Along with this, we must recall that error growth may cause low-dimensionality (Kuhl et al. 2007), and that low dimensional regions are more amenable to forecast improvements, such as with the addition of dropsondes (Oczkowski et al. 2005).

Looking at Figure 18, in terms of the dimensionality of the variables just 6 h into the simulation, θ has the highest domain-averaged E-dimension in each of the three ensembles. This means that in our simulations, θ spreads variance into more error directions more quickly than does wind and water vapor. Also, since there are 14 members in each of these ensembles, it means that 13 is the natural limit of the E-dimension. Note, however, that in both the intervariable and interensemble comparisons (Figure 18), 12 appears to be the limit of dimensionality using these types of ensemble methods and the WRF model. What this means essentially is that there are just slightly fewer significant ensemble error variance directions than there are ensemble members. Considering the relatively simple method of implementation of the breeding method, this high number of error variance directions shows that a quite complicated relationship between the ensemble members arises quite quickly in principal variables.

While θ appears to have slightly higher E-dimension values than both vapor and

wind, the E-dimension of these latter two are quite similar. These comparisons hold not only 6 h into the simulation (Figure 18), but also when plotting domain-averaged E-dimension over time (not shown). This likely relates to the fact that θ errors grow notably more quickly than water vapor and u/v values between rescalings (Table 4) during the entire 24 h cycling period. Note that we saying the error growth *during the breeding period* will increase error dimensionality afterwards, because the errors are being accounted for in the finalized perturbations, but as mentioned by Kuhl et al. (2007), error growth during the simulations themselves will reduce E-dimension. While an explicit orthogonalization among the perturbations of the breeding methods would make this even more the case, this relationship between E-dimension after the breeding period and error growth during the breeding period makes sense, as the method is constructed such that the fully-bred perturbations already contain cumulative error growth in many different directions. This is a unique feature of the breeding method and likely creates the higher E-dimension for FNLq3 and FNLth3 compared with CNTL near the start of the simulation. Ensemble errors occurring during the subsequent *free portion* of the simulation, however, often create areas of *low dimensionality (*Kuhl et al. 2007). In fact, during breeding, the BGM method is so effective at creating directions of variance that the subsequent ET scheme, which attempts to create orthogonal perturbations with respect to an inverse analysis error variance norm (Wei et al. 2008) and on which the control ensemble is based, was instead more particularly focused on ensuring that its variance directions capture the space in which the forecast error evolves.

When comparing E-dimension and spread in Figure 17, the domain averaging done masks much in terms of spread's spatial development for each of the ensembles. To better observe the effect of the various perturbation techniques of interest, we have calculated the *local* 850 hPa water vapor spread for each of the three ensembles of

interest over the entire simulation period. This is plotted in Figures 19, 20, and 21 for the CNTL, FNLq3, and FNLth3 ensembles, respectively. As seen here, there are indeed many interesting patterns worth noting that are masked when considering the domain-averaged metrics in Figure 17. One of the main items to notice in these three figures is the spatial spread differences among the ensembles at the start of the simulation period. For example, notice the initial (0000 UTC 24th), low-resolution, water vapor biases spread over disparate areas of the domain in the CNTL ensemble (Figure 19) compared with the higher resolution and localized spread of the FNLq3 ensemble (Figure 20) and the expected lack of vapor spread in the FNLth3 ensemble (Figure 21). Looking at Figure 19, in some parts of the domain, the vapor standard deviation among the members reaches 2 g/kg, but these areas are not necessarily collocated with the proto-Ernesto. Comparing the resultant spread among the three ensembles at end of the simulation period (1200 UTC 25th; Figure 19-21), the spatial patterns of the spread between both breeding schemes are quite similar, and notably different from those of the control ensemble. This is interesting in that the FNLq3 and FNLth3 ensembles contain separately bred fields (vapor for FNLq3 and θ for FNLth3), while the rest of their IC fields come from the FNL analysis. The final spread patterns could well be due to the initial spatial pattern similarities between bred θ and vapor at 0000 UTC 24th, as they show areas of high θ collocated with areas of low vapor (Figure 15,16), but it is remarkable that this similarity again arises near 36 h. By 1200 UTC 25th, the control ensemble shows few if any areas over the ocean with a spread less than 0.5 g/kg, while the FNLq3 and FNLth3 ensembles show a broad swathe of such values across the eastern half of the domain. Notice also how well the shape of the wave manifests itself in the spread patterns for FNLq3 and FNLth3, especially.

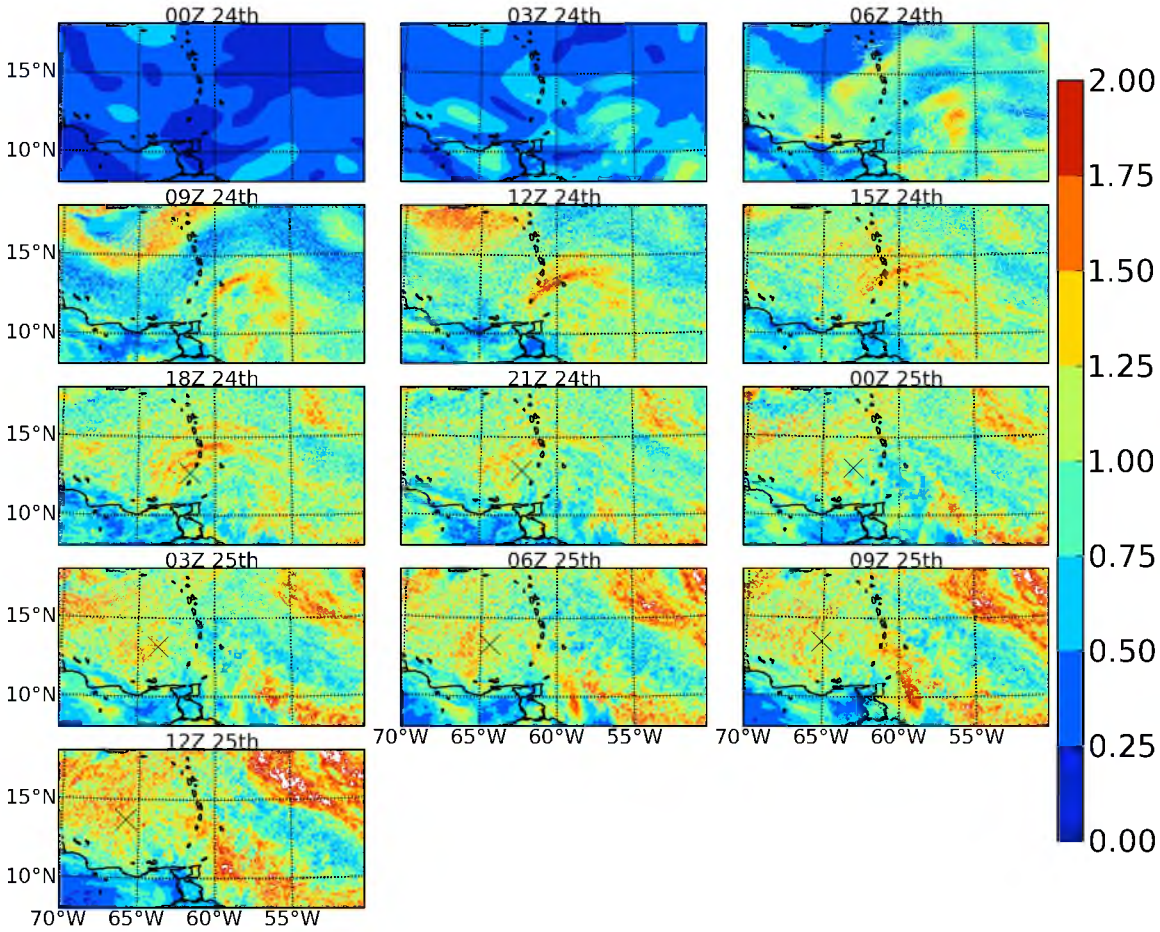

Figure 19. Water vapor (g/kg) spread at 850 hPa over time for CNTL ensemble. Calculation of spread is done as standard deviation with respect to ensemble mean. Black X marks NHC best track TC location after genesis.

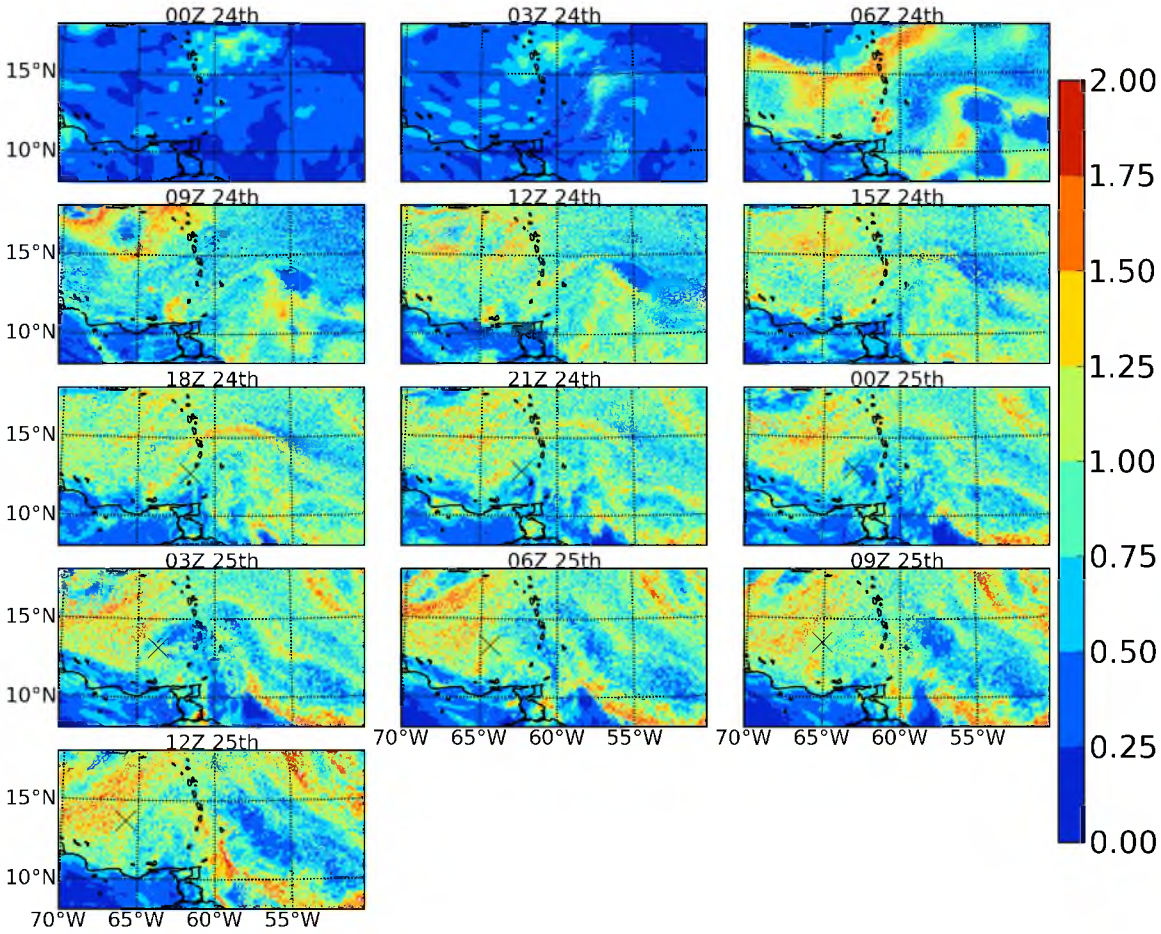

Figure 20. Water vapor (g/kg) spread at 850 hPa over time for FNLq3 ensemble. Calculation of spread is done as standard deviation with respect to ensemble mean. Black X marks NHC best track TC location after genesis.

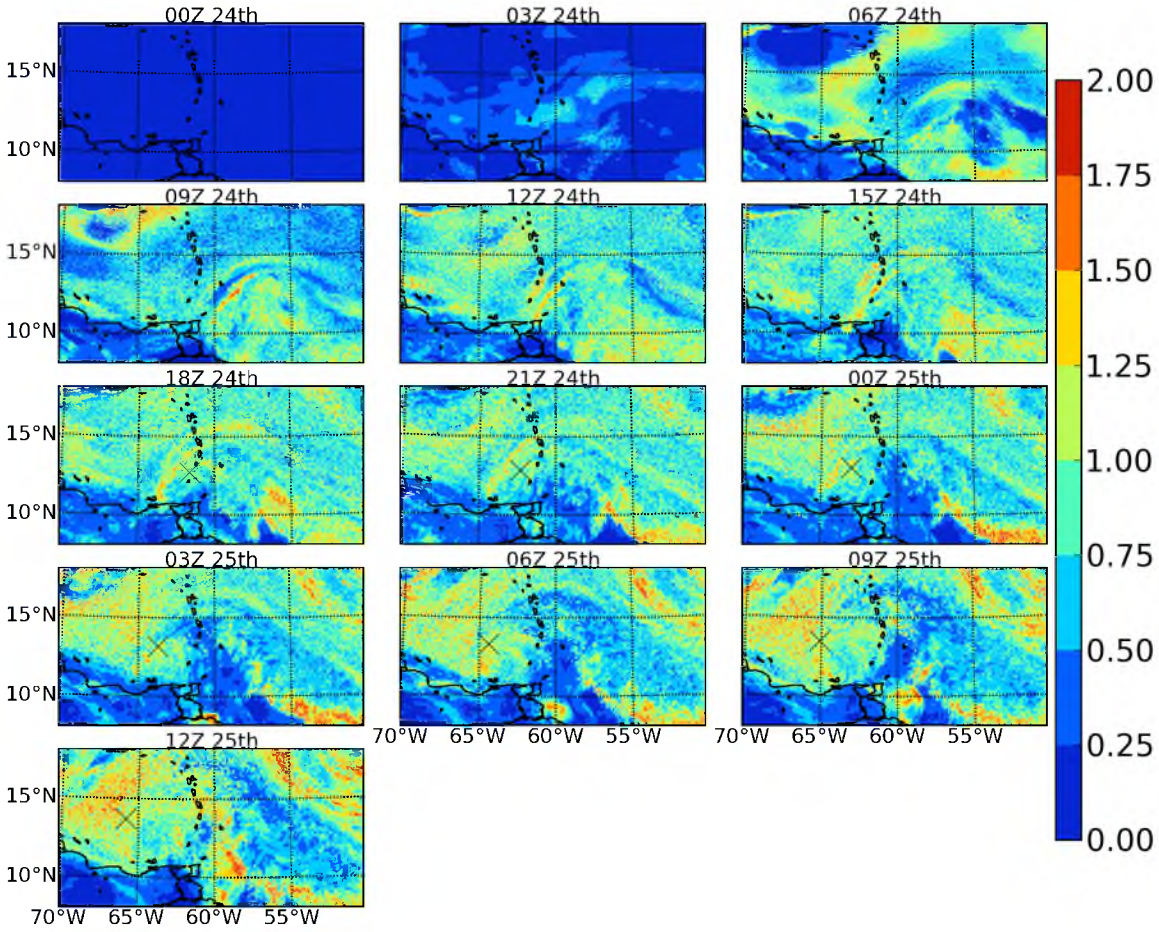

Figure 21. Water vapor (g/kg) spread at 850 hPa over time for FNLth3 ensemble. Calculation of spread is done as standard deviation with respect to ensemble mean. Black X marks NHC best track TC location after genesis.

How E-dimension and spread change by radius from TC center

The high local variability seen in the previous E-dimension and spread calculations was notable to the extent that we now examine the relationship between these metrics at various distances from the forecast TC representing Ernesto. To do this, we used a binning technique to cull only those values at predetermined distances from the relevant TC center over the 36 h of the simulation period. While the domain average used in Figure 17 masked interesting local areas of interest, the data presented in Figure 22 attempt to provide a TC-relevant view of the spread and E-dimension behavior over time, as now we only use water vapor data within 200 km radius of the TC center at 850 hPa. That is, we average the values from each of the 10x10 grid point areas that fall within 200 km of the simulated TC center. This way we can easily determine the differences between domain-scale and localized ensemble spread and error imensionality. First, note how E-dimension and spread have much higher variability over time near the TCs (Figure 22) than on the large scale in general (Figure 17), likely due to the enhanced error growth occurring there (Oczkowski et al. 2005) compared to larger radii. Note also that from 3-12 h into the simulation, the near-TC spread and, especially, E-dimension values are notably higher than they are over most of the domain (Figure 17). What we find generally is that there is a much closer correlation between E-dimension and spread in bred ensembles compared with the control ensemble. Notice the relatively low-dimensionality beginning around 1200 UTC for all ensembles, which reverses around 27 h into the simulation. For the FNLq3 and FNLth3 ensembles, this pattern is closely followed by the values of spread. This U-shaped pattern of low E-dimension from 1200 UTC 24th to 1200 UTC 25th appears similar to that of Figure 2 in Oczkowski et al. (2005), except that the pattern in our data is compressed in time. Comparing this pattern against Figure 17, we notice that the low-dimensionality around 24 h into the simulation

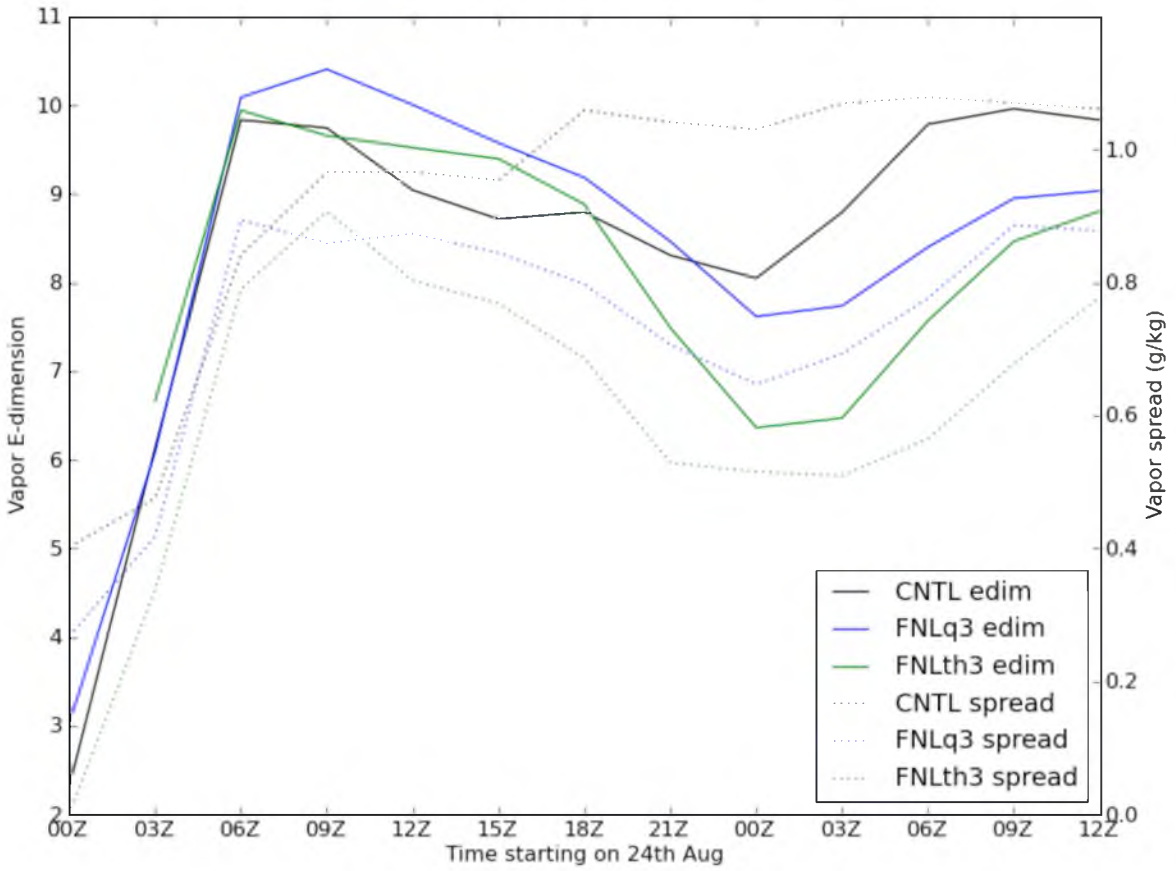

Figure 22. Time series of error E-dimension and spread (g/kg) for 850 hPa water vapor within 200 km radius of simulated TC center in the CNTL, FNLq3, and FNLth3 ensembles.

is much more severe near the TCs than away from them, which is what we would expect considering they are regions of high baroclinicity compared to the Tropics in general. And, as Oczkowski et al. (2005) and Kuhl et al. (2007) have noted, areas of high baroclinic energy conversion are especially prone to low-dimensional behavior.

In terms of the interensemble comparisons, the breeding ensembles show higher E-dimension growth, especially for FNLq3, compared with the control ensemble in the first 12-15 h of the simulation. After 2100 UTC 24th, and with these specific breeding configurations, there is no consistent pattern seen between these or the CNTL ensemble in terms of the degrees of ensemble error variance, or E-dimension. Spread, on the other hand, does appear to be consistently lower for the breeding compared with the control ensemble, and well demonstrates the effect of perturbing one versus several fields (CEFS or CNTL) when creating tropical ensembles.

In order to investigate the significance of the relationship between E-dimension and spread over the simulation period, a bootstrapped correlation coefficient is calculated for the control, FNLq3, and FNLth3 ensembles. First, we must note that this bootstrap method draws 1000 data samples of the E-dimension and spread vectors within 200 km radius of the TC center, for each ensemble, and computes the correlation statistics for each. We perform a bootstrap for the correlation coefficient to provide some idea of confidence limits, which are important here because there are only 13 data points (1 every 3 hours) for both of the metrics we are correlating. For each of the three ensembles, we calculate a cumulative distribution function of the bootstrapped correlations and present it in Figure 23.

First, note how the majority of the data points show that the CNTL ensemble has a vapor E-dimension and spread correlation of 0.7-0.9. At the same time, there is a small

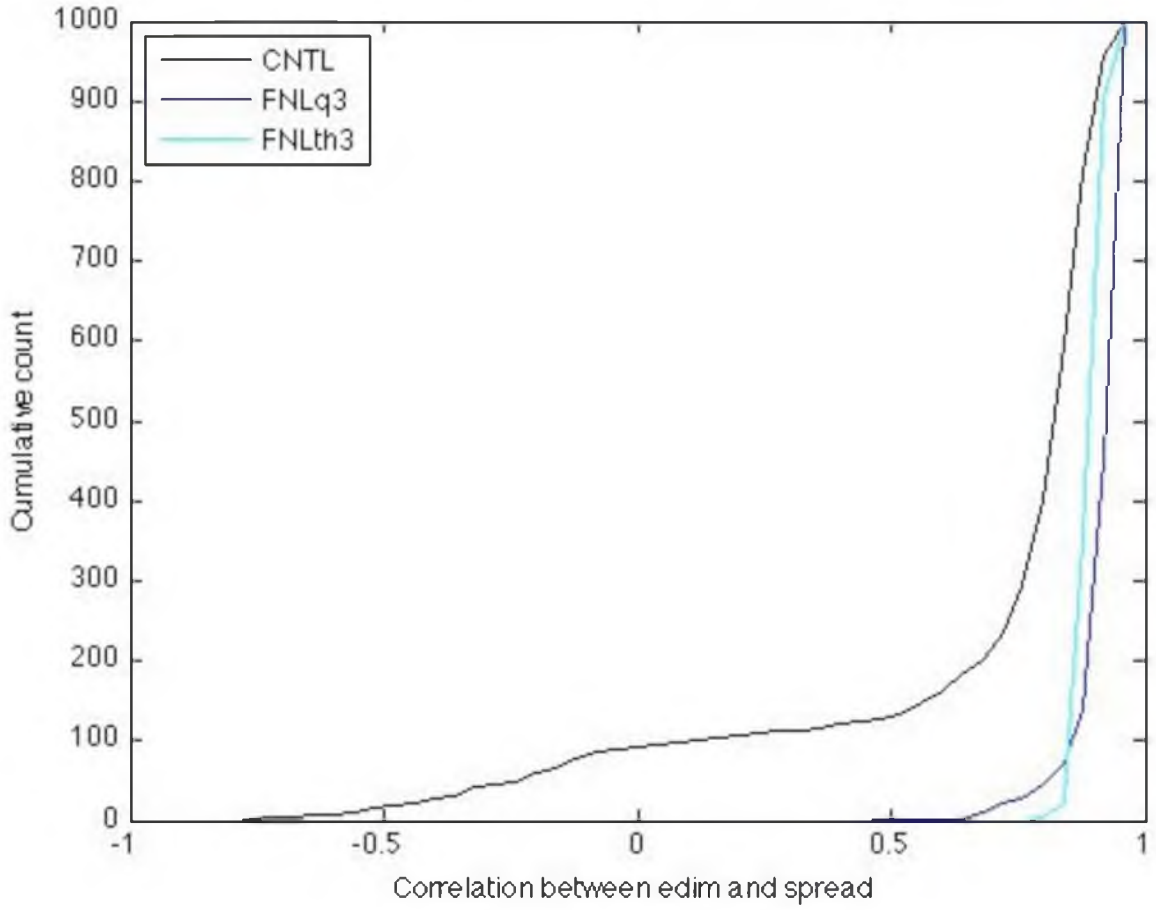

Figure 23. Cumulative distribution of bootstrapped correlation coefficient between E-dimension and spread based on 850 hPa water vapor within 200 km radius of simulated TC center over the entire simulation period of CNTL, FNLq3, and FNLth3 ensembles. The two sets of values for each ensemble were bootstrapped for 1000 samples.

but notable amount of sample constructs which calculate the CNTL correlation coefficient as anywhere from -0.7 to 0.7. This, of course, demonstrates that there is a notable amount of uncertainty in the correlation calculation for the control ensemble. Compare this result with FNLq3, whose spread-dimension correlation values are much more highly concentrated near 1, with only a small number of constructs providing correlation values from 0.5 to 0.8. Thus, compared with the control ensemble, the vapor breeding ensemble has a much stronger relationship between the spread of vapor and its error dimensionality. FNLth3 is similar to FNLq3, wherein the majority of the calculations place the coefficient near 0.9; in this case, however, almost no data constructs calculate the correlation as being below 0.8. Thus, breeding leads to a higher correlation between lower level water vapor spread and error dimensionality compared to CNTL. While breeding θ and water vapor leads to a very similar average correlation value, however, there is notably less uncertainty in the measure when breeding θ (Figure 23).

While previous metrics were based on E-dimension and spread values within 200 km radius of the TC center, it is instructive to view the way that the error dimensionality of several key variables changes by radii from the TC. Using a binning technique for each of the three ensembles in question, E-dimension for water vapor, θ, and wind is calculated over time at particular increments from the corresponding TC center (Figure 24). This demonstrates that E-dimension near TCs is higher than that far from TCs mainly from 0000 to 1800 UTC. As stated earlier, E-dimension values near the TC decay quite rapidly after 12 h, which is what we would predict based on the studies by Oczkowski et al. (2005) and Kuhl et al. (2007). These authors noted that baroclinic energy conversion leads to areas of low dimensionality. That this process would be exacerbated near TCs after model spin up not only makes sense, but is quantified in

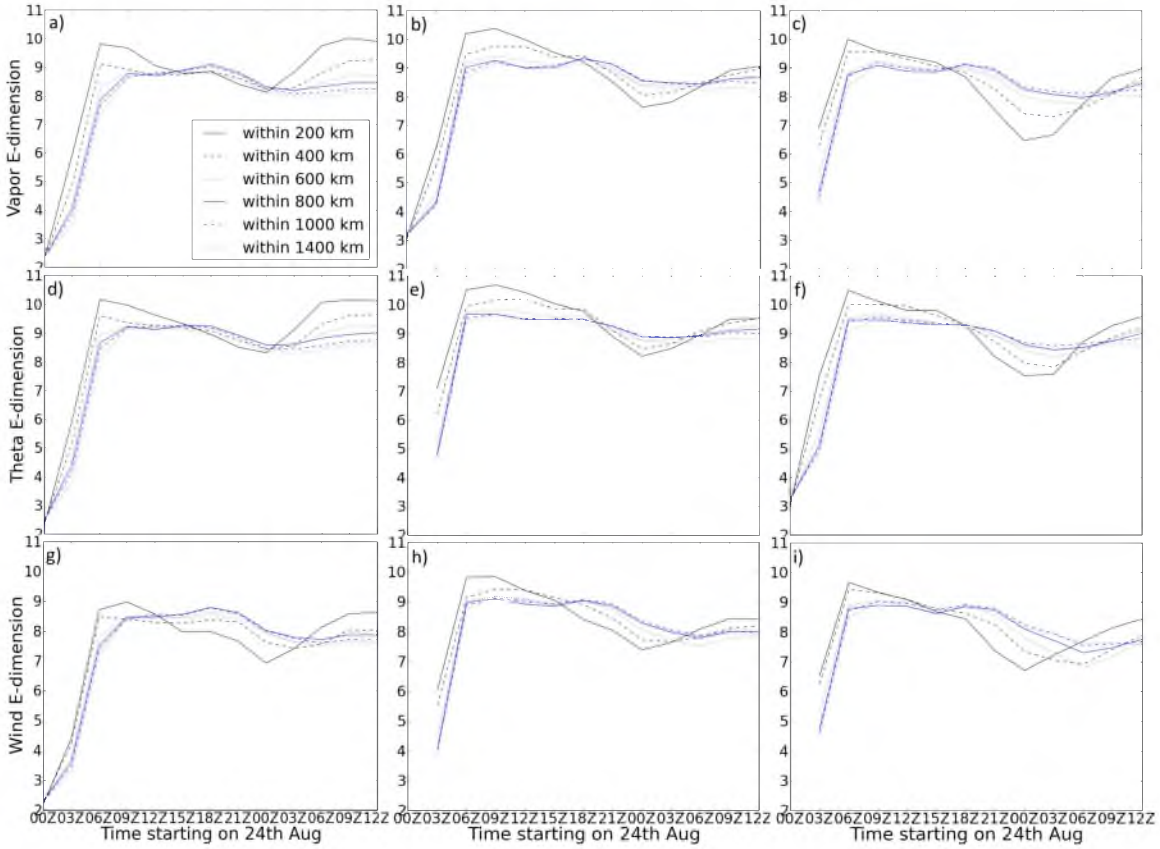

Figure 24. Time series of error E-dimension at 850 hPa and by radii from the simulated TC center for (a-c) water vapor, (d-f) theta, and (g-i) wind. The first column shows results for the CNTL ensemble, second column shows the FNLq3 ensemble, and the third column shows the FNLth3 ensemble.

Figure 24. This near-TC "U" shape was noted by Oczkowski et al. (2005), wherein they explained that E-dimension decreases only in the initial phase of model integration, reaches a minimum from 12 to 48 h hours "depending on the atmospheric flow," and then increases with time. The U shape in Figure 2 of Oczkowski et al. (2005) was only focused on midlatitude variability. The fact that, in our study, we bin E-dimension by distance from the TC helps us to quantitatively determine how tropical dimensionality becomes more like that of the midlatitudes near TC environments.

In terms of interensemble comparisons, and considering vapor dimensionality, first notice that the U shape over time is steepest, when considering 200 km average values, in terms of the FNLth3 (Figure 24c) ensemble, and is shallowest in the control ensemble. This would indicate that θ breeding might lead to more baroclinic error growth and energy conversion near the TCs of interest than would an ensemble based on vapor breeding. The fact that the FNLth3 ensemble would have the most near-TC error growth would make sense, as the TCs in this ensemble are notably more intense than those of the FNLq3 and CNTL ensembles. Of course, more intense TCs are accompanied by stronger local regions of baroclinicity compared to weaker storms.

The E-dimension of the θ (wind) fields over time, and at various distances from the TC center, are shown in Figure 24d-f (g-i). Note that vapor and θ E-dimension patterns are similar from 3 to 9 h. Large differences emerge later, however, when comparing the values across radii. For example, notice that at 0000 UTC 25th, the small to large-radii differences are significantly higher for vapor (Figure 24b,c) than they are for θ (Figure 24e,f). While these differences are most notable in FNLth3, they are present in both breeding schemes. These small to large-radii dimensional differences between θ and vapor may be attributed to error growth differences *by radii* being larger for vapor compared with that of θ.

In terms of possible explanations of CNTL vs BGM differences, as Oczkowski et al. (2005) noted, the typical decrease in forecast dimensionality is caused by baroclinic and barotropic energy conversion, divergence of the ageostrophic geopotential flux, and eddy kinetic energy transport. Thus, it is likely that the lack of a large, near-TC, dimensionality decrease in the control ensemble is due to a lack of one of these things in the near-TC region compared with the other ensembles. The strength of the FNLth3 TCs compared to those of CNTL and FNLq3 fit well with the mechanisms explained by Oczkowski et al. (2005) and our E-dimension results. Besides these midsimulation differences, we also note that, at all radii, wind dimensionality values are consistently lower at 0600 UTC 24th compared with those of vapor and θ. These differences occur for all three ensembles. This could be caused by wind field ICs that contain fewer perturbation dimensions compared with what occurred in the initial vapor and θ fields. To be clear, while E-dimension earlier was calculated for the derived wind, the values in this table are for the fundamental model variables u and v. Looking at Table 4, it appears that reduced error growth rates (especially in the U component) compared to θ and vapor, during the 24 h breeding period, are likely the cause of this reduced dimensionality after the simulation starts at 0000 UTC 24th. Recall, that error growth during the breeding period *appears* to have the opposite effect on error dimensionality from 0000 to 0600 UTC than does error growth which comes *after* the breeding period, for reasons explained earlier. This line of thinking is supported by the fact that E-dimension values at 0600 UTC are also consistently higher for θ compared with vapor, and θ had notably faster error growth during the breeding period than did vapor.

The fact that the low-dimensional behavior starts so abruptly in most fields around 9 h into the simulation is cause for further examination. Notice that the phenomenon, even though it is most strongly present near the simulated Ernesto, is significant enough

to cause a decrease in the domain-averaged E-dimension vapor values after 1800 UTC, starting roughly 9 h after the near-TC dimensionality decrease. What factor might be causing the abrupt drop in significant error variance directions? First, there is a strong warming occurring in these ensembles at 850 hPa after 1500 UTC (Figure 5b). This notable increase in domain-averaged θ may be due to heightened baroclinic and convective energy conversion occurring near the TC. These types of energy conversions, again, are known for creating areas of low-dimensionality. This fits especially well, as Figure 24 is based on E-dimension calculated at 850 hPa and that this is near the height at which latent heating would be occurring due to TC strengthening; note that this same warming is occurring at 700 hPa (Figure 6b). Hohenegger and Schar (2007) explained that mesoscale error growth can be particularly strong in regions of convective instability, which may explain these notable midsimulation drops in error dimensionality at small radii. In addition, while baroclinic processes are likely taking place near the tropical disturbance throughout the simulation (thus working to decrease E-dimension), during the start of the forecast period, the fact that spin-up is occurring and establishing high-resolution patterns and flow in the inner-most domain likely causes the E-dimension of most fields to increase overall. Once this dimensionality increase due to spin-up is removed, the effects of baroclinic and convective energy conversion could easily cause the over all decline in E-dimension values we see after 0900 UTC in Figure 24.

Spectral analysis of ensemble error and perturbations

It has long been documented that spatial spectra of atmospheric fields follow fairly consistent patterns within the mesoscale (Blumen 1978; Nastrom and Gage 1985). For example, the latter study used data from over 6000 commercial aircraft flights and found that both the temperature and wind spectra consistently showed a spatial spectrum with a

slope of near $k^{-5/3}$ for wavelengths below a few hundred km and a slope of k^{-3} for wavelengths from 1000-300 km. Thus, a slope of k^{-3} was roughly observed at the synoptic scale with that of $k^{-5/3}$ roughly in the mesoscale (Nastrom and Gage 1985; Rotunno and Snyder 2008). This slope essentially describes the power-law behavior of atmospheric energy spectra, K^{p}, where K is wavenumber magnitude and p is the slope of the spectrum. In fact, Hohenegger and Schar (2007) looked at this issue, and further clarified that "convective and baroclinic instabilities live in the $k^{-5/3}$ and k^{-3} spectral regimes, respectively." Subsequently Berner et al. (2009) found that large-scale models in general had a difficult time reproducing the $k^{-5/3}$ slope found in nature and developed a perturbation method to not only help resolve this issue, but also reduce ensemble underdispersion overall.

In terms of our study, these spectral slopes can be used to gauge whether, and under which circumstances, the error growth in the various ensembles is more convective or baroclinic. In addition, these high-resolution ensembles can be compared as to which best produce the $k^{-5/3}$ slope below roughly 400 km and the k^{-3} slope at larger wavelengths. Plotted in Figure 25 are these spectral results for our three ensembles of interest at 3, 6, and 9 h into the simulation period. This calculation is done by first subtracting the analysis from the 500 hPa geopotential height ensemble mean, performing a Fast Fourier Transform on the data in the meridional direction, and then taking a zonal mean of this spatial spectral data. Analyzing the data in this way not only allows us to view the behavior of the various ensembles in terms of the slopes mentioned above, but, more importantly, also allows us to determine whether, and at which wavelengths, certain ensembles provide spread to match their error (see both Figure 25 and 26). What we find is that, as expected, there is much more spectral energy on the large scales (i.e., at long

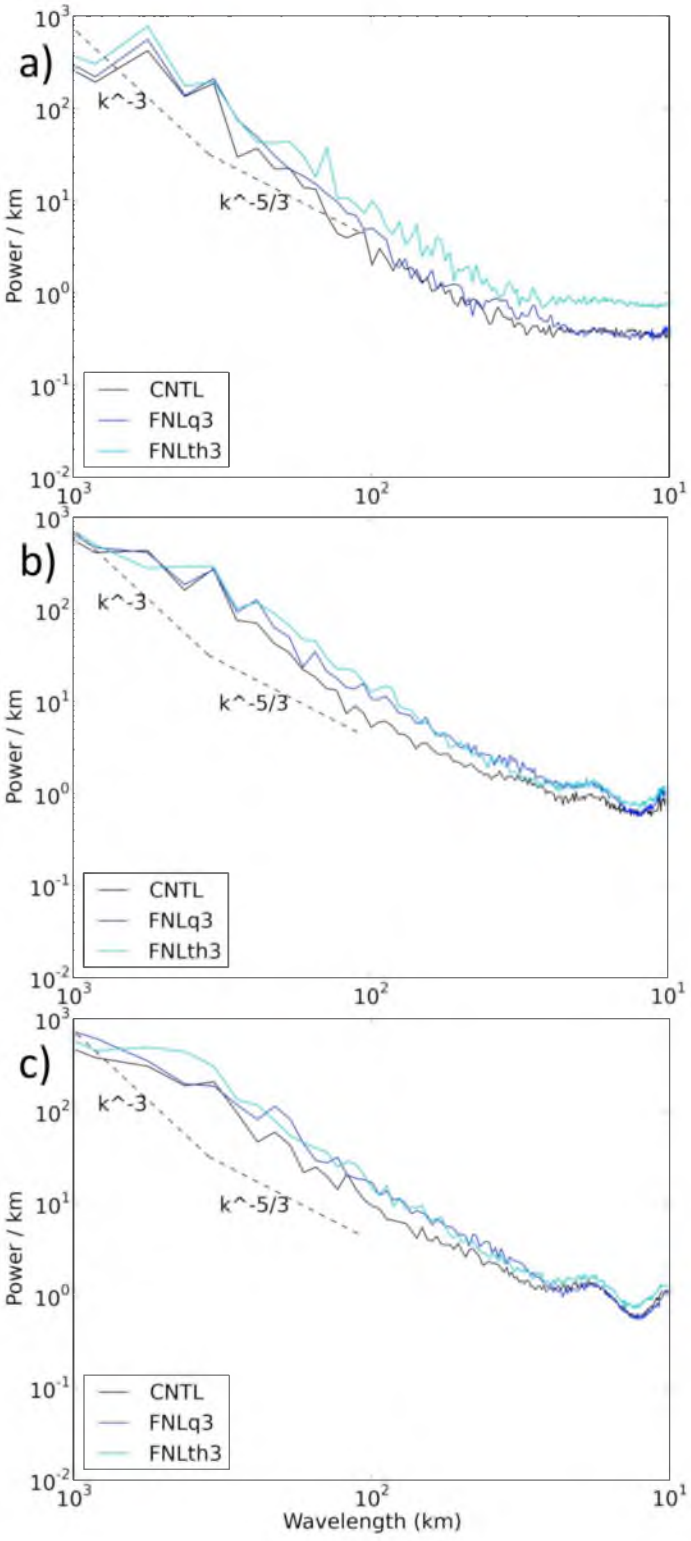

Figure 25. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height error across a zonal and ensemble average for CNTL, FNLq3, and FNLth3 at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

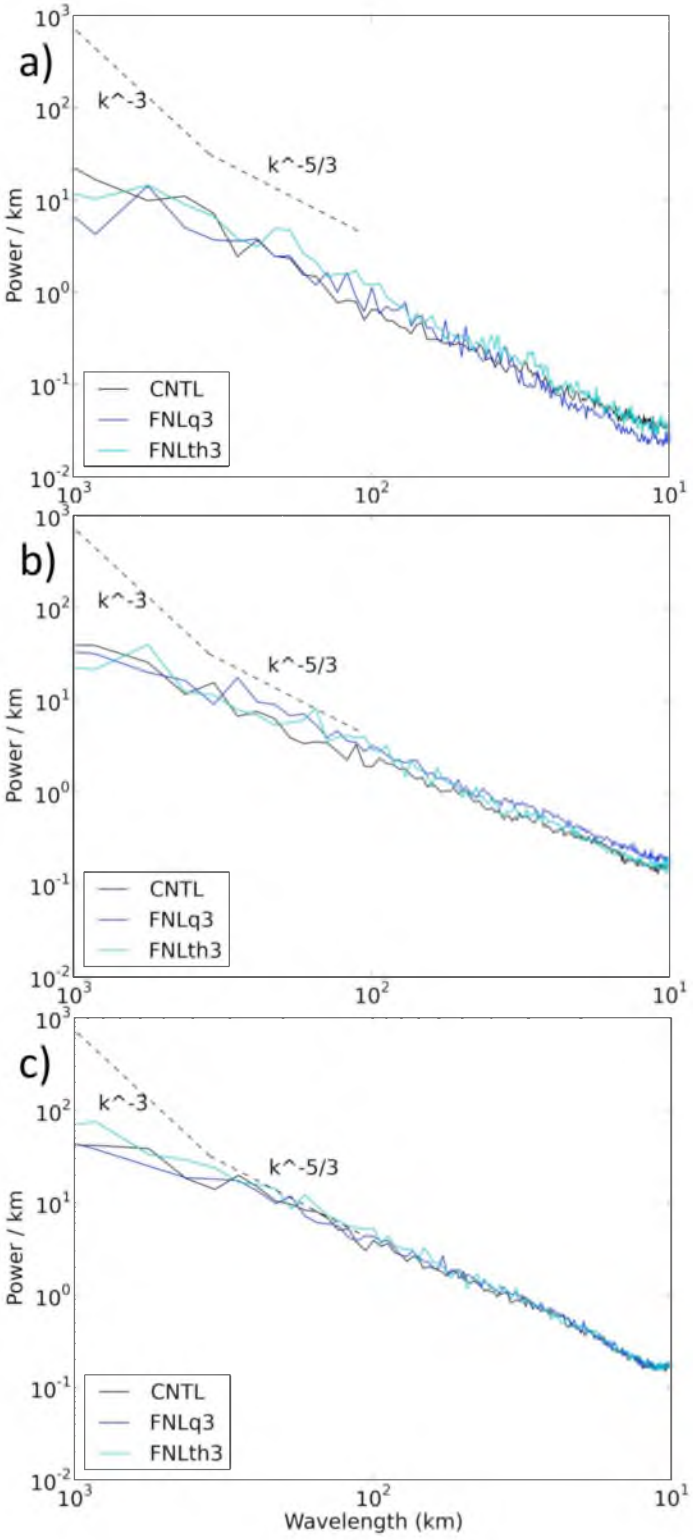

Figure 26. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height perturbation from the ensemble mean, then across a zonal and ensemble average, for CNTL, FNLq3, and FNLth3 at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

wavelengths). The fact that this energy decreases toward the small scales provides the spectral slopes which have been discussed. Note that the ideal $k^{-5/3}$ and k^{-3} spectral slopes in their respective regimes are placed in the figure and that below roughly 25 km, the numerical damping of the model produces unphysical spectral results. Three hours into the simulation, we note that there is significantly more error in the FNLth3 ensemble, across much of the spectrum, compared with the FNLq3 and CNTL ensembles. Note that by 0600 UTC, this quickly changes and the CNTL shows notably less error from 50 to 200 km length scales, compared to both breeding ensembles. Note that while this gap narrows by 0900 UTC and 1200 UTC (not shown), there is still a slight less energy in the CNTL ensemble in terms of error around 100 km wavelength compared to the other ensembles. Whether this CNTL ensemble also enjoys a more favorable error-spread relationship is discussed below.

While this absolute amount of spectral error energy among the ensembles is notable in Figure 25, the other major item is the difference in spectral slopes. Remember that most models fail to produce the transition, found in nature, from a spectral slope of $k^{-5/3}$ below 400 km to k^{-3} above 400 km (Berner et al. 2009). While it is difficult to compare such slopes across the spectrum in Figure 25, there are some useful interensemble observations to be made. First, at large length scales, the slopes for each of the ensembles are quite similar, although none produce slopes of k^{-3} because of the size of our domain. Because of this, we will generally focus on the ability of the ensembles to produce the $k^{-5/3}$ slope at the high-resolution end of the spectrum. By 0900 UTC, we find that the three ensembles show slopes just slightly steeper than the $k^{-5/3}$ found in nature. Since the CNTL ensemble shows slightly lower error particularly around 100 km, this ensemble appears to show a slope slightly steeper than $k^{-5/3}$ at larger wavelengths and slightly less steep than that at smaller wavelengths.

As ensembles are generally noted for being underdispersive (Berner et al. 2009), and due to the fact that we have now examined error spectra, it is appropriate to analyze the corresponding perturbation spectra. These are presented in Figure 26 and are calculated by first defining the perturbation at each time as each member minus the ensemble mean, again in terms of 500 hPa geopotential height. Subsequently, the mean of the perturbations across the ensemble is calculated, a Fast Fourier Transform is performed, and a zonal mean is taken. These are plotted over the first 9 h of the simulation and for the three ensembles being studied. First, one of the most notable differences between Figure 26 and 25 is the fact that the perturbation energy is about an order of magnitude lower than is the error energy, which is the underdispersion Berner et al. (2009) lamented. While this difference occurs at most wavelengths, there are important differences across the spectrum. For instance, observe the way the error spectra show a concave shape at 0300 UTC, whereas that of the perturbation spectra is convex; for reference, note that the theoretical k slopes are similarly placed in Figure 25 and 26. The ensembles in question are thus especially underdispersive at large and small length scales, as there is a particular lack of energy in the perturbation spectra, compared with that of the error, at large and small wavelengths. While the perturbation energy at 100 km wavelength is lower than that of the error, the differences between these two in this particular part of the spectrum are relatively insignificant.

In terms of the perturbation spatial energy spectra differences among the ensembles (Figure 26), it first must be noted that these are markedly less than the corresponding error spectra differences. Note that at 0300 UTC (Figure 26a) the difference are slight, but that the FNLq3 ensemble has less energy than FNLth3 and the control ensemble at length scales greater (smaller) than 300 km (40 km). As with error spectra, the differences between ensembles in terms of perturbation spectra shrink notably from 0300

to 0600 UTC. By 0600 UTC, the bred ensembles show sightly higher spread, especially around a few hundred km wavelength, compared with the CNTL ensemble. Considering the common ensemble underdispersion problem, this increase in FNLq3 and FNLth3 spread is interesting in itself, however, we must recall that it is accompanied by higher error at the same time (Figure 25b). By 0900 UTC (Figure 26c), all three ensembles produce very similar perturbation spatial spectral patterns at almost all length scales. Compare this with the progression of the corresponding error energy (Figure 25), wherein at 0900 UTC (Figure 25c) and 1200 UTC (not shown) the bred ensembles still showed notable differences from the control ensemble, mostly around 100 km wavelength. Essentially, in terms of energy spectra, differences between bred and CNTL ensembles diminish quickly in terms of perturbations but those in terms of error abide for at least 12 hours, which means after 0300 UTC, *the CNTL ensemble provides less underdispersion than the bred ensembles*.

Highlights of Chapter 3

- Θ errors consistently grow faster than those of water vapor. Θ error growth Rates plateau around 12 h into the breeding period, whereas vapor error growth is still increasing (albeit only slowly) 18 h into the period.
- For breeding techniques in the Tropics, a total breeding period length of less than 24 h would provide a sufficient number of cycles such that error growth rates plateau. Eighteen hours would appear to be a long enough period.
- Both BGM and ET breeding methods produce more spatial differences, between members, in terms of the θ rather than vapor. Likely, it relates to higher θ error growth compared to that of vapor.
- FNLq3 ensemble has notable dry bias, compared with FNLth3, that appears to

have a negative impact on TC genesis forecasts.

- Similar to Oczkowski et al. (2005), we find notable drops in error variance dimensionality through the middle of the simulation period for all ensembles. Also, this drop appears exacerbated due to baroclinic error growth and instability surrounding forecast TCs.
- In terms of ensemble spectral differences, we find that the least amount of underdispersion occurs near 100 km wavelength and for CNTL in general.

CHAPTER 4

EFFECTS OF CYCLING PERIOD AND LOCALIZATION

Ensembles with multiple variables bred and varying cycling periods

While the first bred ensemble comparisons focused on the differences resulting from vapor versus θ initial perturbations, the next set focuses on the effect of varying the cycling period during the 24 h breeding interval. This test is performed by using a 3, 6, and 12 h cycling period while breeding wind components u and v, in addition to vapor, and θ. In addition, this is done with an invariant scaling factor, which was similarly used in McLay et al. (2008) and in our previous breeding ensembles. At the same time, these new ensembles, named FNLuvqt3, FNLuvqt6, FNLuvqt12 (Table 2), will provide data on potential differences arising from breeding one variable versus several, as they will be compared against FNLq3 and FNLth3.

Effects of cycling period on error growth

We will begin the analysis of this new experiment with a look at error behavior during the 24 h breeding period. Scaling factors every 3 h for the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles are presented in Table 5. Again, note that relatively large error growth leads to a smaller scaling factor. First, notice the significant differences between the error growth of the different variables. Just as we saw between the FNLq3 and FNLth3 cases, during the entire breeding period, θ errors grow notably faster than do those of vapor. The wind scaling factors typically experience error growth closer to that

Table 5

Comparison of scaling factors for member 1 of

FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles during 24 h breeding period

Time	3 hr cycle				6 hr cycle				12 hr cycle			
	U	V	Q	TH	U	V	Q	TH	U	V	Q	TH
03Z	0.96	0.97	0.83	0.80								
06Z	0.83	0.87	0.75	0.67	0.82	0.85	0.69	0.63				
09Z	0.84	0.80	0.77	0.64								
12Z	0.75	0.68	0.76	0.54	0.74	0.68	0.68	0.51	0.64	0.60	0.60	0.45
15Z	0.80	0.68	0.73	0.58								
18Z	0.74	0.66	0.66	0.58	0.74	0.66	0.58	0.53				
21Z	0.73	0.67	0.73	0.59								
00Z	0.71	0.65	0.77	0.62	0.74	0.67	0.66	0.54	0.59	0.55	0.55	0.44

of vapor than that of θ, and, interestingly, we find that the v wind component generally experiences faster error growth than does the u component. In terms of error growth changes over time, it appears that growth rates plateau from 12 to 1800 UTC. This certainly appears to be the case for at least the 3 and 6 h ensembles. In terms of differences between the cycling periods, fairly consistently we see that the longer breeding cycles produce error growth higher than the shorter breeding cycles, which is to be expected. Nevertheless, the changes do not occur linearly by cycling period. For example, comparing vapor error growth at 1200 UTC it is seen that by changing from 3 to 6 to 12 h cycle periods one changes the scaling factor from 0.76 to 0.68 to 0.60, respectively. In other words, the error growth goes from 32% over 3 h to 67% over 12 h, which is not close to the multiple of four one would expect considering the time difference, prima facie. Doubling the cycling period from 3 to 6 h causes vapor error growth ending at 1200 UTC to just rise from 32% to 47%. Thus, a large portion of the vapor error growth occurs over the first few hours of each cycle, and, while the 12 h cycling period does provide more error growth per cycle, with short breeding periods there are more chances of restarting and breeding error in different directions. Also, the 3 h cycles also likely have the benefit that the corresponding error growth is consistently in the linear growth regime, whereas for the 12 h cycles this is unlikely the case. This can be partially be inferred from the fact that error growth over the 12 h cycles is far from four times the error growth that occurs over the 3 h cycles.

Initial condition differences

Now that the breeding error growth itself has been discussed for the FNLuvqt ensembles at 3, 6, and 12 h cycling periods it is prudent to examine the initial conditions after the 24 h breeding period had finished. Of course, not only are we interested here

in the effect of cycling period on the final perturbation product, but also the comparative effect of breeding four variables instead of just θ or vapor. First off, we return to the bias in Figures 4-7 to ascertain the general placement of the FNLuvqt ensembles, in terms of vapor and θ, at the start of the control simulation period compared to both the analysis and other ensembles being tested. Starting at 700 and 850 hPa (Figures 5a and 6a), it appears that at 0000 UTC 24th, the FNLuvqt ensembles possess about 0.25-0.5 g/kg less water vapor than do both the analysis and the FNLth3 ensemble. Note that the FNLq3 is positioned well below zero as well, and closely aligned with the FNLuvqt ensembles. Depending on height, a shorter breeding cycle period in the FNLuvqt ensembles produces either higher or lower amounts of vapor after the 24 h. Considering what we have seen regarding the significant impact of vapor on genesis, we already expect the FNLuvqt ensembles' disturbances to behave more like those of the ensemble where solely vapor is bred instead of that where only θ is bred. This general clumping in terms of bias does align with experimental design, as the FNLth3 ensemble does not breed or perturb vapor and this would have a bias of zero at 0000 UTC 24th. Similarly, when examining θ bias in the FNLq3 ensemble, it will be zero at the start time as well. Thus, comparisons of θ bias in the FNLuvqt ensembles can only be made against that of the FNLth3 ensemble.

Looking at this metric at 925 hPa (Figure 4), it is evident that the FNLuvqt ensembles align with the FNLth3 ensemble in terms of domain-averaged initial θ. In fact, the breeding of θ by itself places its bias almost exactly in the middle of the FNLuvqt ensemble θ bias at the control start time. It is interesting that the three other variables being bred do not have a notable effect, during the 24 h breeding period, on the ultimate values of either θ or vapor at the end of the period. While all ensembles are similar in terms of initial θ at 850 hPa (Figure 5), at 500 hPa (Figure 7) and especially 700 hPa (Figure 6), the θ breeding causes notable differences between those ensembles and the

analysis by 0000 UTC 24th. Note how at 700 hPa all ensembles breeding θ start the free part of their simulations with domain-averaged θ 0.5 K higher than the rest of the ensembles. Again, whether θ is bred along with other variables does not appear to have a consistent effect on θ bias after the 24 h breeding period. If there is a difference due to breeding cycle *period*, the FNLuvqt6 ensemble appears to produce slightly higher θ values at 0000 UTC 24th (Figures 4-7) at most heights compared to the 3 or 12 h cycling.

Figure 27-29 present the 850 hPa vapor bias of the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles in a plan view for each of their members. In each of these figures, we see an overall slight dry bias compared to the analysis (see also Figure 5). Note the lack of significant spatial differences among the members, similar to what was seen in the control (Figure 13) and FNLq3 ensembles (Figure 15) at this time. There are, however, notable differences in the patterns between these and the FNLq3/th3 ensembles. For instance, while in the FNLuvqt ensembles (Figure 27-29), there is a notable dry bias around 15°N and across much of the domain in the east/west direction, in the control ensemble (Figure 13), the most notable persistent feature is the moist bias throughout the center of the domain and along the western edge from 10 to 15°N. The FNLq3 ensemble (Figure 15) does exhibit the notable dry bias around 15°N similar to the FNLuvqt ensembles, but also shows a significant dry bias in the southeastern corner of the domain, centered on 10°N, which is notably more muted when vapor is bred with θ, u, and v (Figure 27).

Comparing the effect of breeding cycle period on vapor bias spatial structures and magnitudes, it appears that the overall general patterns of moisture bias are similar whether we breed for 3, 6, or 12 h (Figures 27-29); basically, there is a large swath of

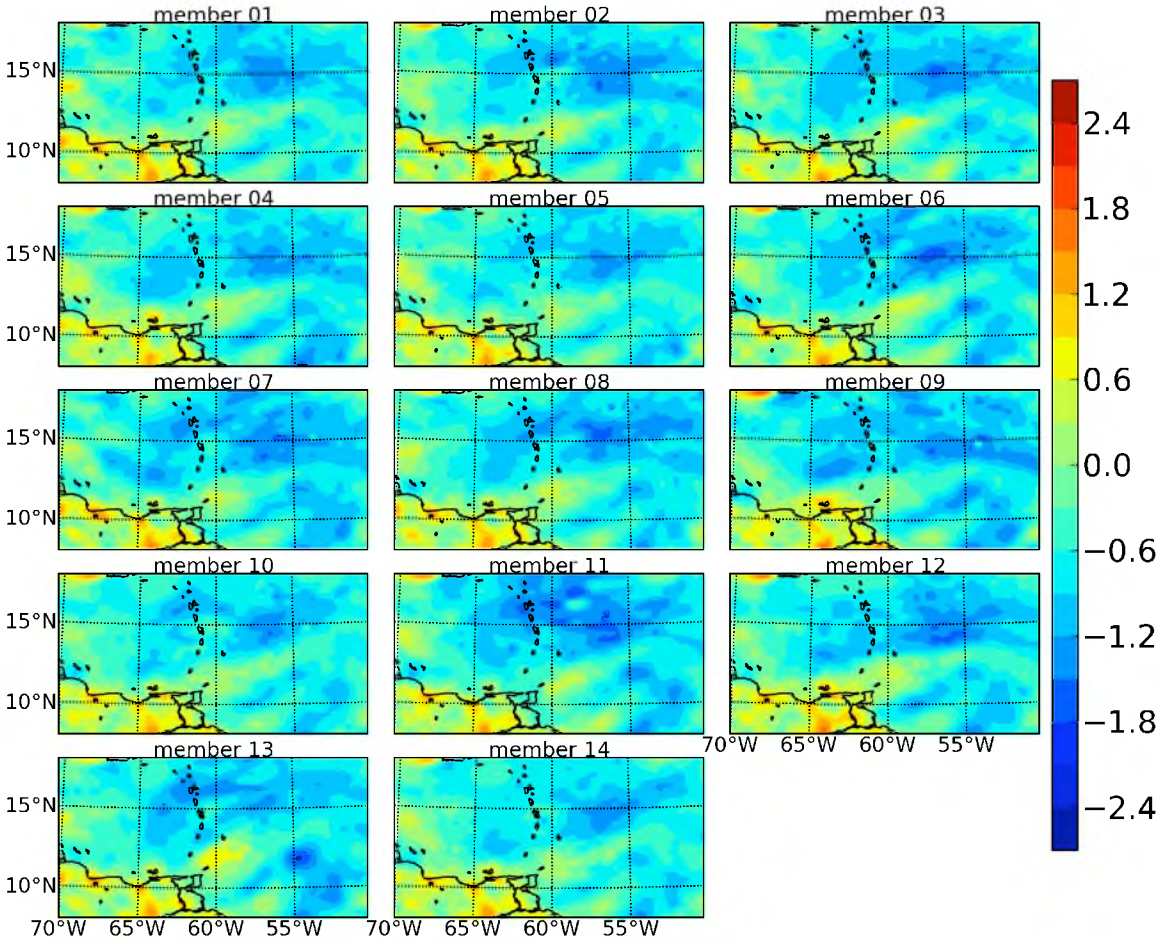

Figure 27. Bias of 850 hPa vapor (g/kg) for each member of FNLuvqt3 ensemble at 0000 UTC 24th (end of breeding period).

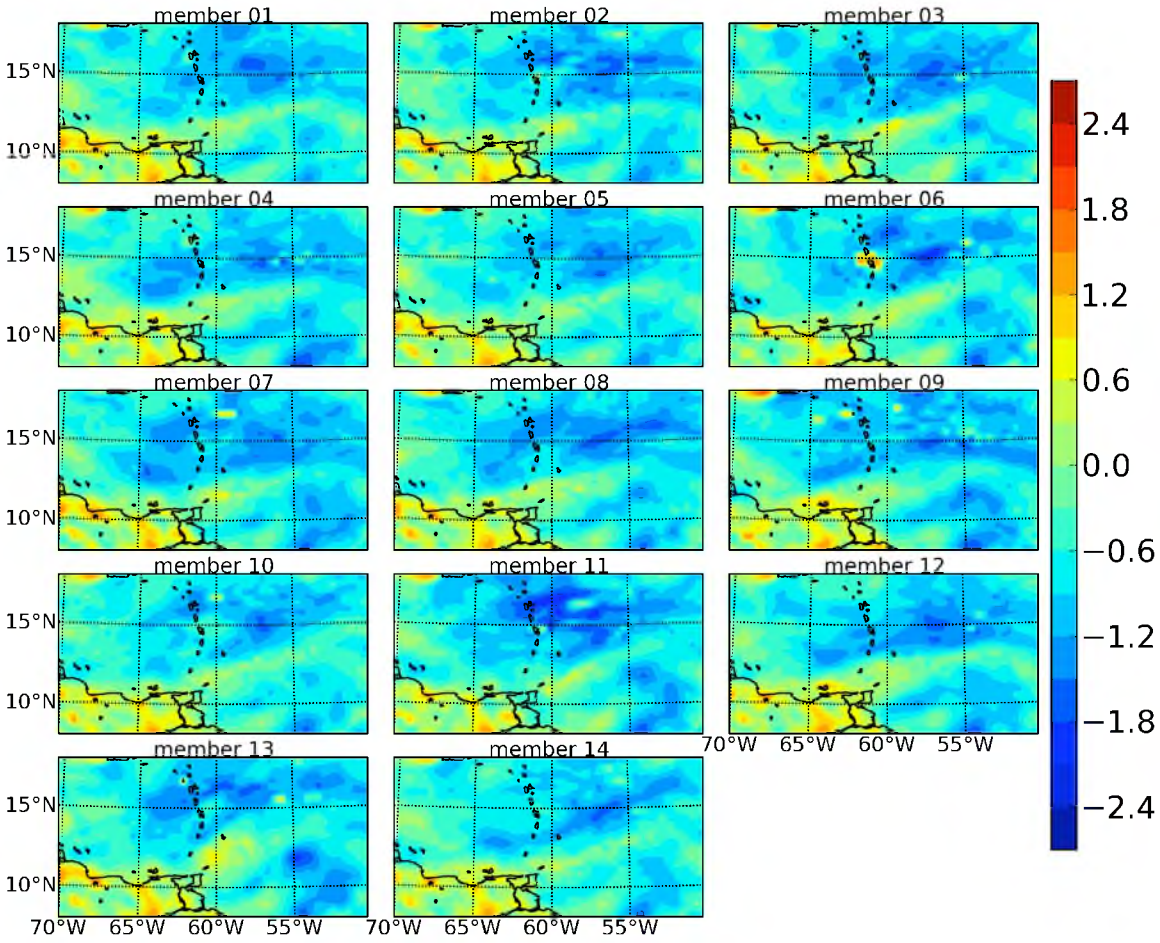

Figure 28. Bias of 850 hPa vapor (g/kg) for each member of FNLuvqt6 ensemble at 0000 UTC 24th (end of breeding period).

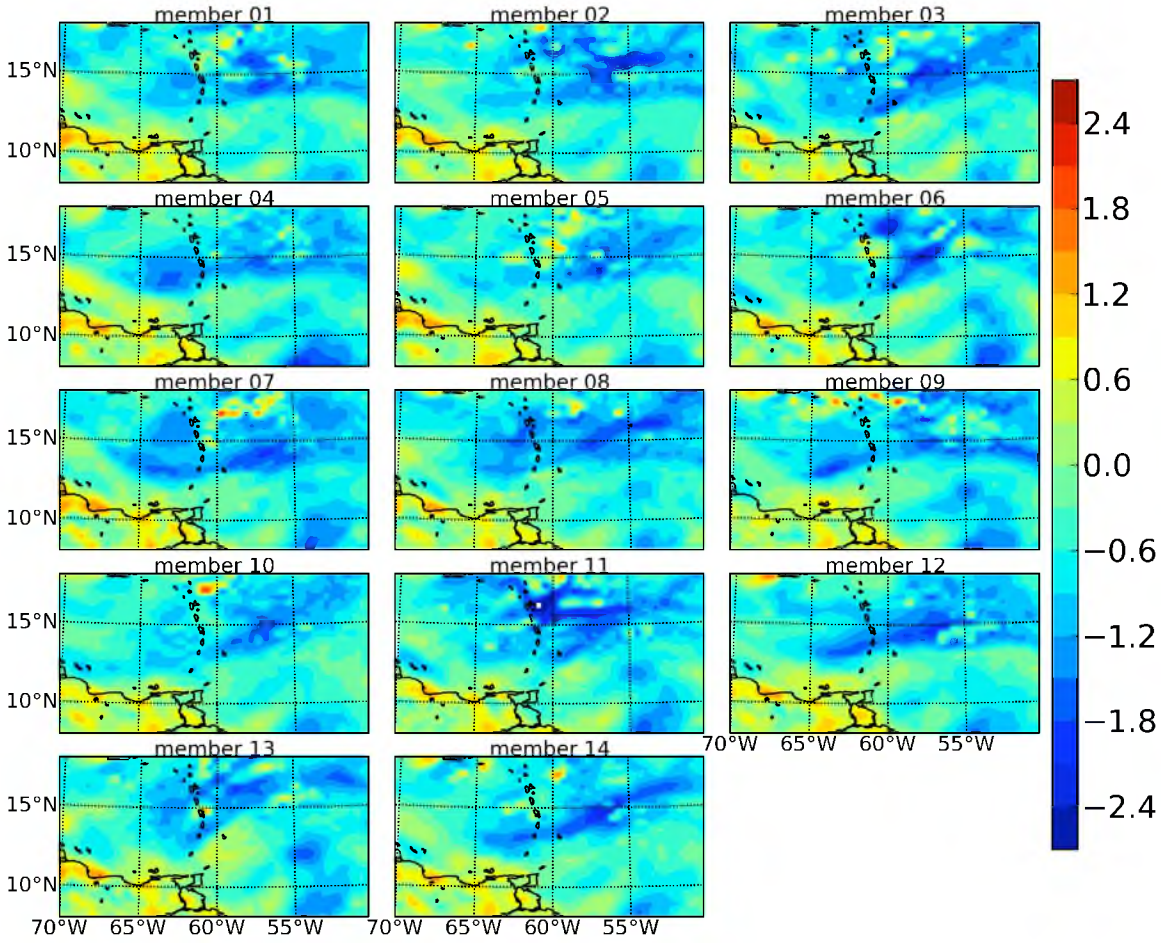

Figure 29. Bias of 850 hPa vapor (g/kg) for each member of FNLuvqt12 ensemble at 0000 UTC 24th (end of breeding period).

dry bias across the northern half of the domains. The small spatial structure differences between the members are much more notable between the FNLuvqt12 members than the FNLuvqt3 members, likely because the longer cycles produce finer scale error patterns.

Of course, there is a tradeoff, which is found in the other notable vapor difference between the ensembles, viz. the magnitudes of these biases in the various regions consistently change with cycling period. For example, note how in the large dry bias across the northern half of the domain, the FNLuvqt6 ensemble produces drier values than does the FNLuvqt3 ensemble, and that the FNLuvqt12 ensemble produces the driest values of all in this region. Focus on almost any of the members and compare across breeding cycle period and one notices that the 12 h cycle causes water vapor bias values to minimize roughly 0.5 g/kg closer to zero than in the 3 h breeding ensemble. Overall, this three ensemble set provides an example of the fine-scale error pattern versus frequency tradeoffs associated with breeding cycle period, and provides a basis with which to compare genesis occurrence, ensemble dimensionality, spread, and error spatial structure over time.

Effects of cycling period on forecasts of TC development

Next, we turn to the actual ability of each of these FNLuvqt ensembles to spin up a vortex and enable TC genesis. Plotted in Figures 30-32 are 850 hPa geopotential heights and vorticity for each of the members of the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles at 1800 UTC 24th August. First, comparing FNLuvqt3 (Figure 30) to FNLq3 (Figure 10) and FNLth3 (Figure 11), we see that the FNLuvqt3 ensemble produces slightly weaker disturbances than the FNLq3 ensemble and notably weaker disturbances than the FNLth3 ensemble. At the same time, the FNLuvqt3 ensemble provides more

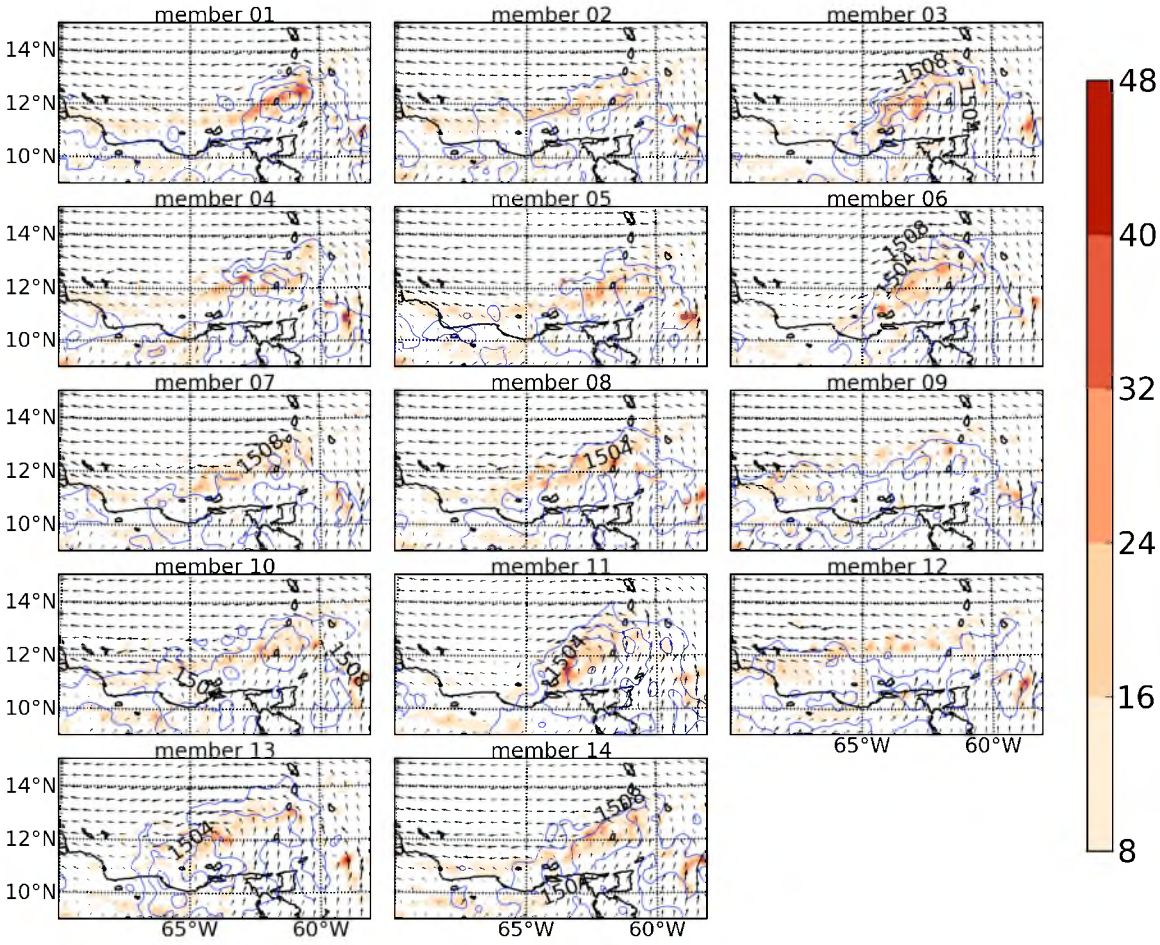

Figure 30. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt3 ensemble members at 850 hPa at 1800 UTC 24th.

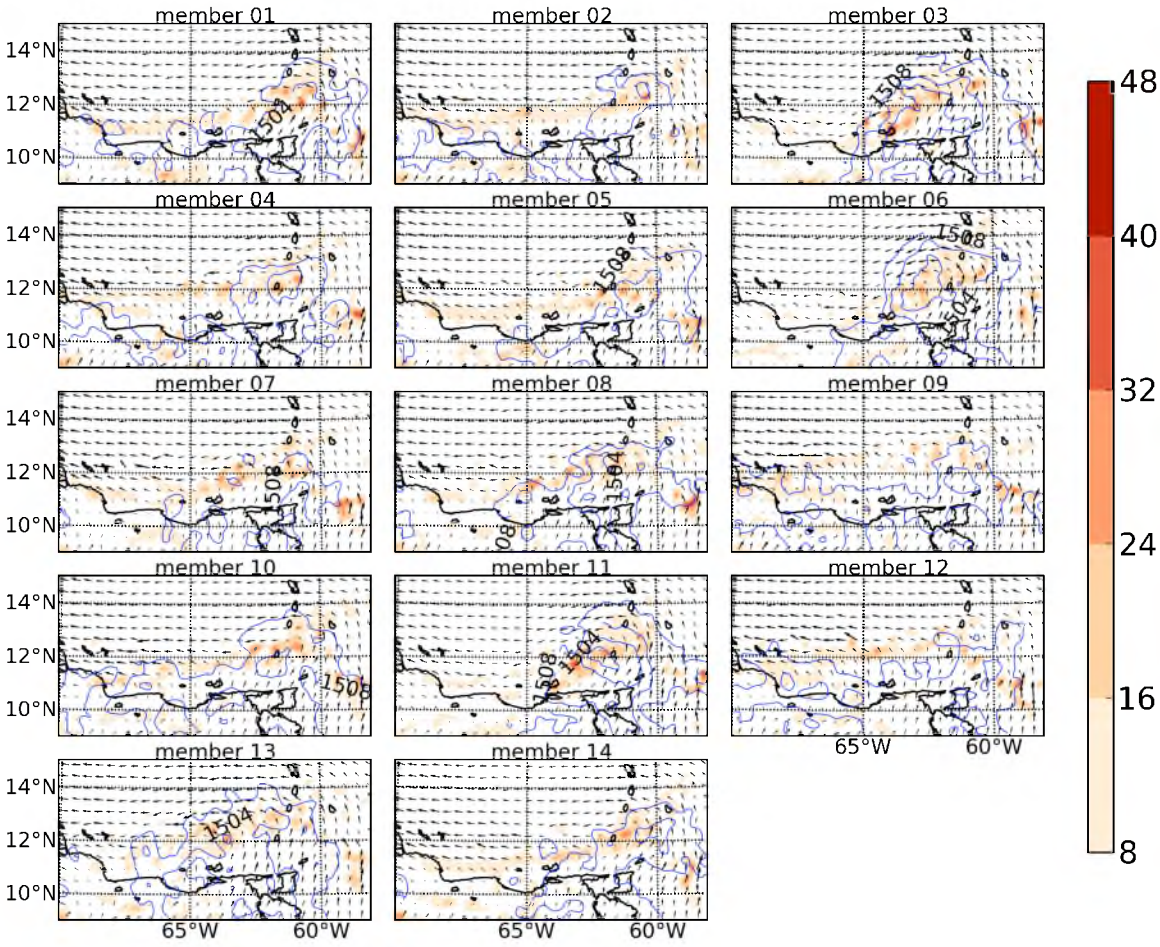

Figure 31. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt6 ensemble members at 850 hPa at 1800 UTC 24th.

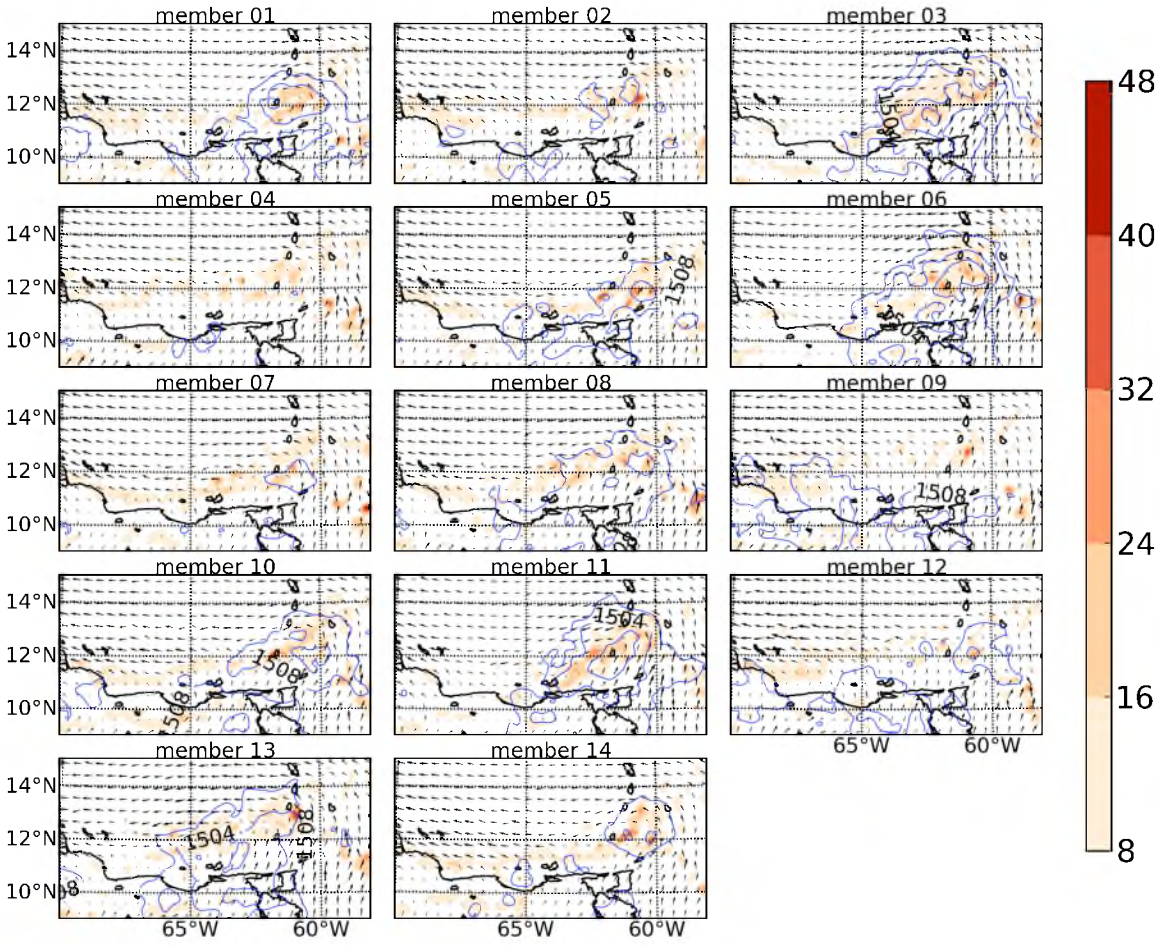

Figure 32. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt12 ensemble members at 850 hPa at 1800 UTC 24th.

variability in terms of both intensity and track, due to u and v being bred and perturbed, compared with both the FNLq3 and FNLth3 ensembles. These two comparisons also are valid for the FNLuvqt6 and FNLuvqt12 ensembles. As was anticipated, the FNLuvqt ensembles likely suffer in terms of their potential to forecast genesis because of their notably dry bias values at 1800 UTC 24th at the important levels of 850 and 700 hPa (Figures 5 and 6). This likely explains the weakness of the TCs at 18000 UTC 24th produced by the FNLq3 ensembles as well. This does not explain, however, why the TCs in the FNLuvqt ensembles are weaker than those in the FNLq3 (Figure 10) ensemble. Part of this explanation might relate to the fact that perturbations to the u and v fields, in addition to the vapor or θ field, might require more time for vortex spinup and the associated TC processes to properly begin. Model balance does not appear to be a large factor, as tests using high-frequency output did not reveal any nonphysical features or waves as the simulations begn.

In terms of the impact breeding cycle period has on TC genesis occurrence, it appears that the shorter breeding period produces stronger TCs. Now, this does not occur for all members, but, generally, the disturbances in FNLuvqt3 (Figure 30) are more intense than those in FNLuvqt12 (Figure 32). The FNLuvqt3 members have smaller initial perturbations at 0000 UTC 24th compared to the FNLuvqt6 and FNLuvqt12 ensemble members and these larger perturbations in some areas may make it harder for TC spinup to occur, as the nonbred FNL fields align with and influence the bred fields. Also note the fact that while the shorter breeding cycles produce stronger TCs, this does not mean that the FNLuvqt3 ensemble produces many more genesis cases compared to FNLuvqt6 and FNLuvqt12 by 1800 UTC 24th. On the contrary, as seen in Table 6, 18 h into the simulation, there is only 1 genesis case in each of the three ensembles. At the end of the simulation period, however, the effect of cycling period is seen more

Table 6

Genesis statistics for

Hurricane Ernesto (forecast began 0000 UTC 24th)

Ensemble experiments	18Z 24th	12Z 25th
FNLuvqt3	01/14	12/14
FNLuvqt6	01/14	09/14
FNLuvqt12	01/14	08/14

clearly, in that the 3 h breeding cycle produces 12 genesis cases compared to the 8 produced by the twelve hour cycling.

How the E-dimension-spread relationship is affected by the cycling period

From genesis implications we now turn toward the relationship between cycling period length and the dimensionality and spread of the ensemble. Plotted in Figure 33 is the E-dimension values 6 h into the simulation period over most of the domain for vapor, θ, and wind at 850 hPa in the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles. Contrary to what we find in Figure 18, the differences between our current ensembles under consideration are not large. In general, we see that for all three cycling periods, E-dimension values across much of the domain for vapor, θ, and wind are already near 10-11. Comparing against Figure 18a, it appears that breeding with four variables does produce higher E-dimension more quickly than does the control ensemble. In terms of the results of breeding one variable versus four, it appears that the latter method produces less dimensional variability over the ocean in terms of θ and vapor. More so than in Figure 18, 6 h into the simulation, FNLuvqt breeding produces notably higher wind E-dimension spatial heterogeneity compared with θ and vapor at 850 hPa. At this early stage of the simulation, this means that lower-level wind dimensionality in general grows with less consistency than does that of θ and vapor. All else being equal, this leads us to believe that in Ernesto's environment, early dropsonde observations of wind fields would more quickly lead to early forecast improvements compared to those of θ or water vapor (Oczkowski et al. 2005). In terms of differences between cycling periods, differences over the ocean for vapor, θ, and wind dimensionality are quite small at this early stage of the simulation.

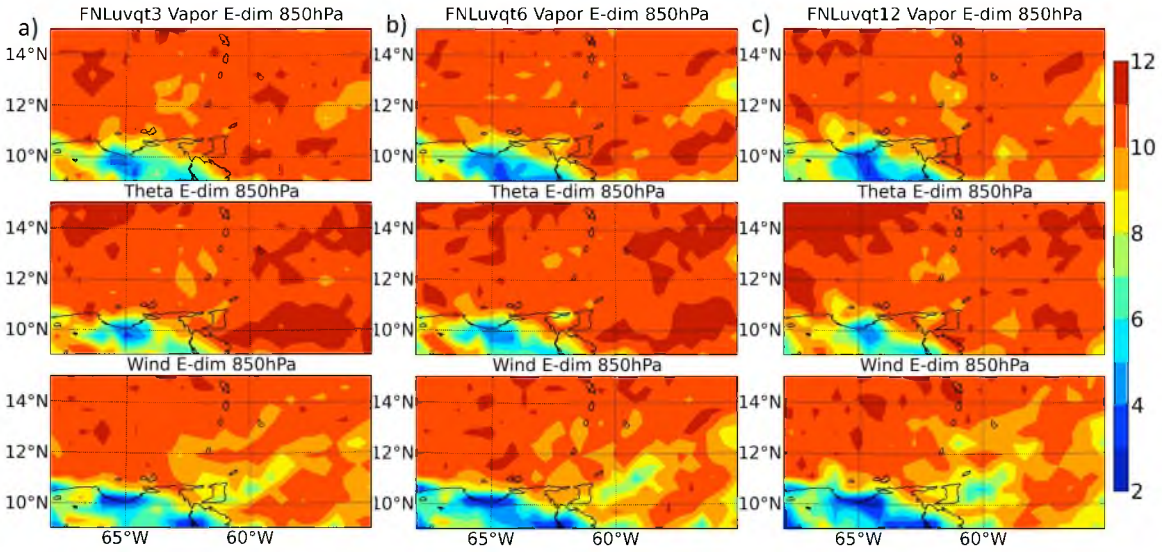

Figure 33. Error E-dimension for water vapor, theta, and wind at 850 hPa at 0600 UTC 24th August for the a) FNLuvqt3, b) FNLuvqt6, and c) FNLuvqt12 ensembles.

To better discern the effects of cycling period on ensemble dimensionality at each stage of the simulation, the E-dimension is averaged within 200 km radius of the TC centers over time and plotted in Figure 34. Similar to Figure 22, this is done at 850 hPa in terms of water vapor and plotted alongside ensemble vapor spread. While there is a notable difference in the rate of E-dimension growth between the FNLuvqt ensembles 3 h into the simulation, the dimensionality of the FNLuvqt ensembles does not appear to vary significantly through the simulations due to changes in cycling period. We do note that dimensionality growth at the start of the simulation is quicker when breeding vapor, θ, u, and v instead of just vapor or θ individually (see Figure 22); by 0600 UTC 24th, the FNLuvqt ensembles are 0.5 higher than the CNTL, FNLq3, and FNLth3 ensembles, while the drop in FNLuvqt3 E-dimension after 18 h is much less than that for either FNLq3 or FNLth3. While we do not have enough data to say that this applies generally, in our ensembles, it appears the interaction between the error growth of disparate variables during their breeding makes it such that the resulting error variance directions in the vapor field are consistently higher than when the vapor or θ field is bred by itself.

The fairly consistent relationship between E-dimension and spread over the simulation period for the FNLq3 and FNLth3 ensembles is quite notable (Figure 22). By contrast, when breeding the four variables together, the vapor E-dimension decay from 9 to 24 h is less sharp and the drop off in vapor spread over the same time period is nonexistent. Instead of vapor spread decreasing after 9 h in the FNLuvqt ensembles, it actually increases rather consistently until the end of the simulation period, with values around 0.2 g/kg higher when compared with FNLq3 and FNLth3. Thus, when breeding the four variables instead of θ or vapor individually, the link between dimensionality and

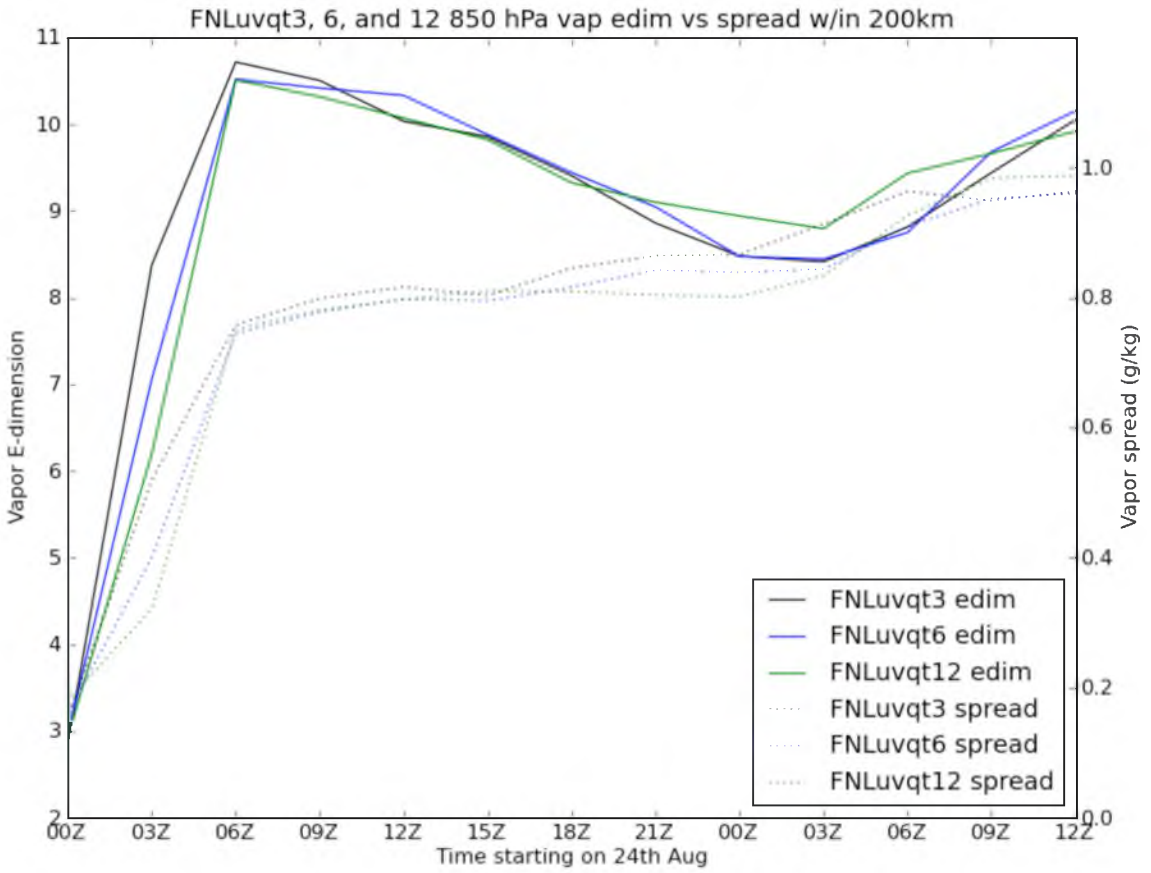

Figure 34. Time series of error E-dimension and spread (g/kg) for 850 hPa water vapor within 200 km radius of simulated TC center in the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles.

spread across the domain appears to break. Notice the increase in spread when breeding multiple variables is much more monotonic compared to when just one variable is bred (Figure 34 versus 22). The FNLuvqt ensembles show vapor spread of around just one g/kg less than the CNTL ensemble (Figure 34 versus 22), which helps quantify the similar impact of ultimately perturbing four key variables (FNLuvqt) versus perturbing many variables (CNTL, based on ET).

In order to evaluate the extent that dimensionality and spread decouples in the FNLuvqt ensembles, the correlation coefficient, for the values within 200 km radius of the TC center, was calculated over the simulation period, with uncertainty estimates being provided by a 1000 sample bootstrap calculation. The result, in cumulative distribution form, is presented in Figure 35 for the FNLuvqt3, FNLuvqt6, and FNLuvqt12 ensembles. Compared against Figure 23, we see that, indeed, breeding multiple variables at once appears to notably decrease the certainty of a high correlation between vapor dimensionality and spread in our forecasts of Ernesto. Interestingly, there are some notable features of Figure 35 that are not evident in the previous figure. For example, we see that the FNLuvqt3 ensemble, compared to 6 and 12 h breeding cycles, has a much lower certainty of providing a high correlation between vapor E-dimension and spread. Note around one-third of the FNLuvqt3 samples falls below a correlation of 0.5. Further, it appears that the FNLuvqt12 ensemble has the most statistically significant high correlation values between vapor E-dimension and spread among the three ensembles. Thus, in our particular breeding setup and forecasts of Ernesto, which is using a spatially invariant rescaling factor, the *longest breeding cycle* provides the strongest relationship between water vapor error dimensionality and spread.

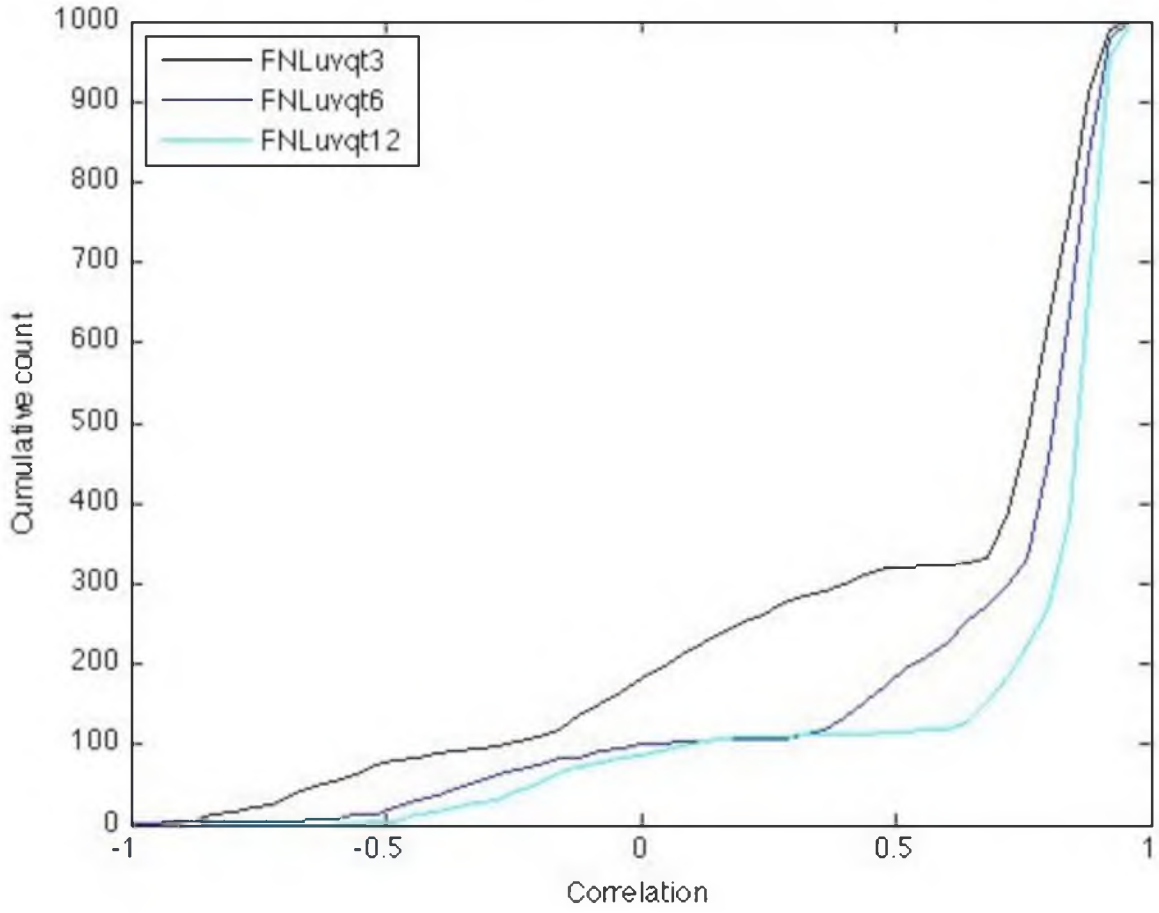

Figure 35. Cumulative distribution of bootstrapped correlation coefficient between E-dimension and spread based on 850 hPa water vapor within 200 km radius of simulated TC center over the entire simulation period of FNLuvqt3, FNLuqt6, and FNLuvqt12 ensembles. The two sets of values for each ensemble were bootstrapped for 1000 samples.

Effect of breeding cycle on dimensionality at various radii from TC center

While the above results are based on data from within 200 km radius of the TC center, it is helpful to study the change in dimensionality at various radii from the storms. Plotted in Figure 36 is 850 hPa E-dimension over time and at various radii for water vapor, θ, and wind in the FNLuvqt ensembles. Compared with Figure 24, we here notice that near-TC dimensionality is less often lower than that at large radii. Similar to the CNTL, FNLq3, and FNLth3 ensembles (Figure 24), here we notice that there is a significant drop in dimensionality after the first 12 h of the simulations for each variable, ensemble, and most radii under consideration, although the drop is less when breeding multiple variables. Again, because these relatively low dimensional environments, after 12 h, are more notable at small versus large radii, they appear to be due to the baroclinic effects related to TC intensification processes; in Figure 36, we notice that this difference-by-radii is more the case for wind compared with θ and vapor.

In terms of the details of the change in θ and vapor error dimensionality by radii, the increase toward the TC is usually monotonic, but seldom uniform over radii. In terms of evidence that changes in dimensionality are driven by TC-relevant processes in particular, we note that it is from 600 to 400 km and also from 400 to 200 km radii that we find the largest changes in E-dimension. These radii are where TC outer rainbands are positioned and denote the sharpest point where we switch from background tropical error growth patterns to those associated with the baroclinicity and convective processes associated with TC-related structures.

In terms of the differences in dimensionality by radii *between* the FNLuvqt ensembles, the maximum dimensionality, changes over time, and changes by radius do not differ much by breeding cycle period. The most notable difference comes in the wind

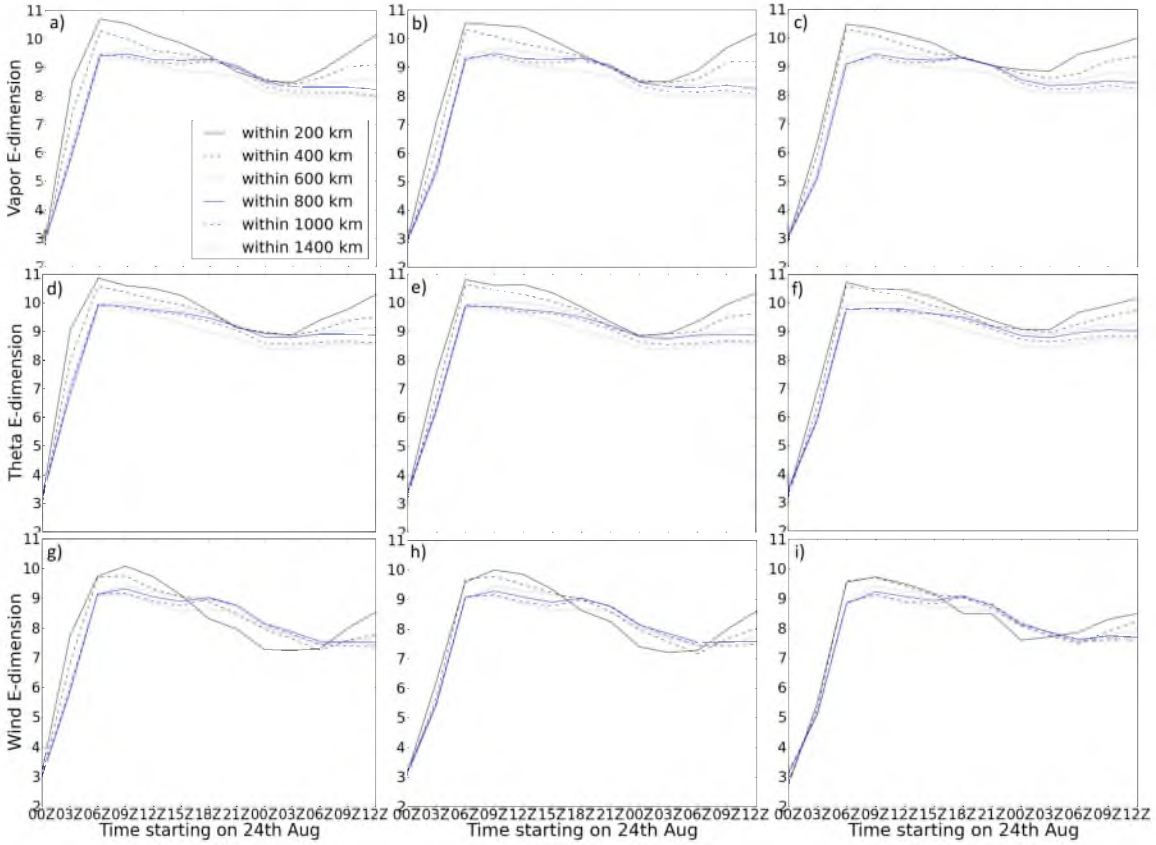

Figure 36. Time series of error E-dimension at 850 hPa and by radii from the simulated TC center for (a-c) water vapor, (d-f) theta, and (g-i) wind. The first column shows results for the FNLuvqt3 ensemble, second column shows the FNLuvqt6 ensemble, and the third column shows the FNLuvqt12 ensemble.

field. Note how there is slightly more near-TC versus large radii wind dimensional variability in the FNLuvqt3 ensemble compared with the FNLuvqt12 ensemble. These differences are not large, but they do lead us to believe that, of the three variables, for unknown reasons, wind dimensionality is most affected when changing breeding cycle period length.

Initial perturbation and error spatial spectra by breeding cycle period

Now that we have looked at the differences between breeding cycles in terms of error growth, genesis occurrence, and ensemble dimensionality; it is appropriate to look at the spectral structure of the errors and how these evolve at the start of the simulation period. Plotted in Figure 37 is the Fast Fourier Transform (FFT) of ensemble-averaged 500 hPa geopotential height error at 0300, 0600, and 0900 UTC. The FFT is calculated across the meridional direction and the resultant spatial spectrum is averaged zonally. First, comparing this figure against Figure 25, we see that changing the breeding cycling period results in much smaller spatial error structure differences compared with changing the variable bred, or when comparing these versus the control ensemble. At 0300 UTC, we see that the FNLuvqt12 ensemble, among the three, has the lowest amount of energy at middle and small wavelengths (10 to 100 km). This is unexpected, considering the relationship between time and length scales in the atmosphere, wherein one would predict the FNLuvqt3 ensemble, with its focus on short-term error growth, would have less error than FNLuvqt6 and FNLuvqt12 at shorter wavelengths particularly. This expectation comes from the close atmospheric relationship between time and scale. Overall, however, the plot succinctly quantifies the differences and shows us that the large-scale error structures are not much affected by changes in cycle period.

In this figure, we plot for reference the k^{-3} and $k^{-5/3}$ theoretical power spectra slopes,

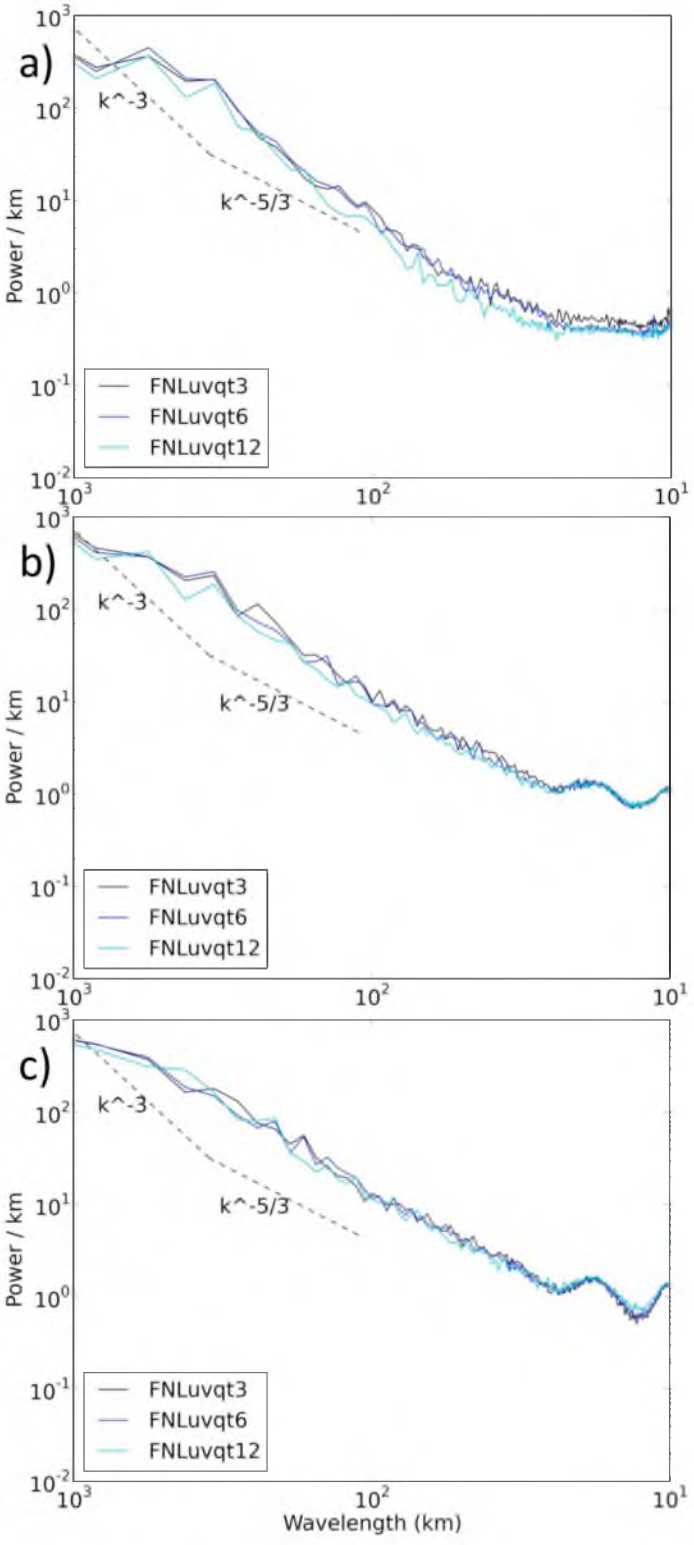

Figure 37. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height error across a zonal and ensemble average for FNLuvqt3, FNLuvqt6, and FNLuvqt12 at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

for their corresponding wavelength bands, that have been found to occur in the atmosphere (Nastrom and Gage 1985). As might be expected considering the fact that we are studying tropical ensembles, we note that after 6 h, across almost the entire spectrum (Figure 37b,c), the data are closely aligned with the $k^{-5/3}$ slope, although very slightly steeper. This indicates that at this time, the error spectrum is characteristic of convective rather than baroclinic processes at these wavelengths (Hohenegger and Schar 2007). By 6 h, the three FNLuvqt ensembles have aligned as well, at nearly all parts of the spectrum. Thus, in terms of error spatial structures, the effect of breeding period is nearly indistinguishable after 6 h of simulation time. In addition, by 6 h the differences between the FNLuvqt and FNLq/th3 ensembles (Figures 25 and 37) are negligible, which indicates that most differences in terms of error spatial structures caused by unique breeding implementations are gone by this time.

When discussing error spatial spectrum pattern, it is illustrative to also view the perturbation spatial spectrum. This latter calculation, which is similar to that for error, except differences are now taken from the ensemble mean instead of the analysis, is performed and presented in Figure 38. This shows perturbation spectrum for the FNLuvqt ensembles over the start of the simulation period. Among the FNLuvqt ensembles, we see the typical perturbation convex pattern across the spectrum, with power per wavelength increasing almost equally for all ensembles over time. Compared with the FNLuvqt ensembles, the FNLq3 and FNLth3 ensembles at 0900 UTC (Figure 26) show slightly higher perturbation energy at wavelengths greater than 100 km. Interestingly, it thus appears that breeding one variable occasionally increases perturbation 500 hPa geopotential height spatial structures compared to the breeding of multiple variables. Looking at Figures 37 and 38 to evaluate this ensemble underdispersion, which is observed in the fact that the error spectra is higher than perturbation spectrum energy, we

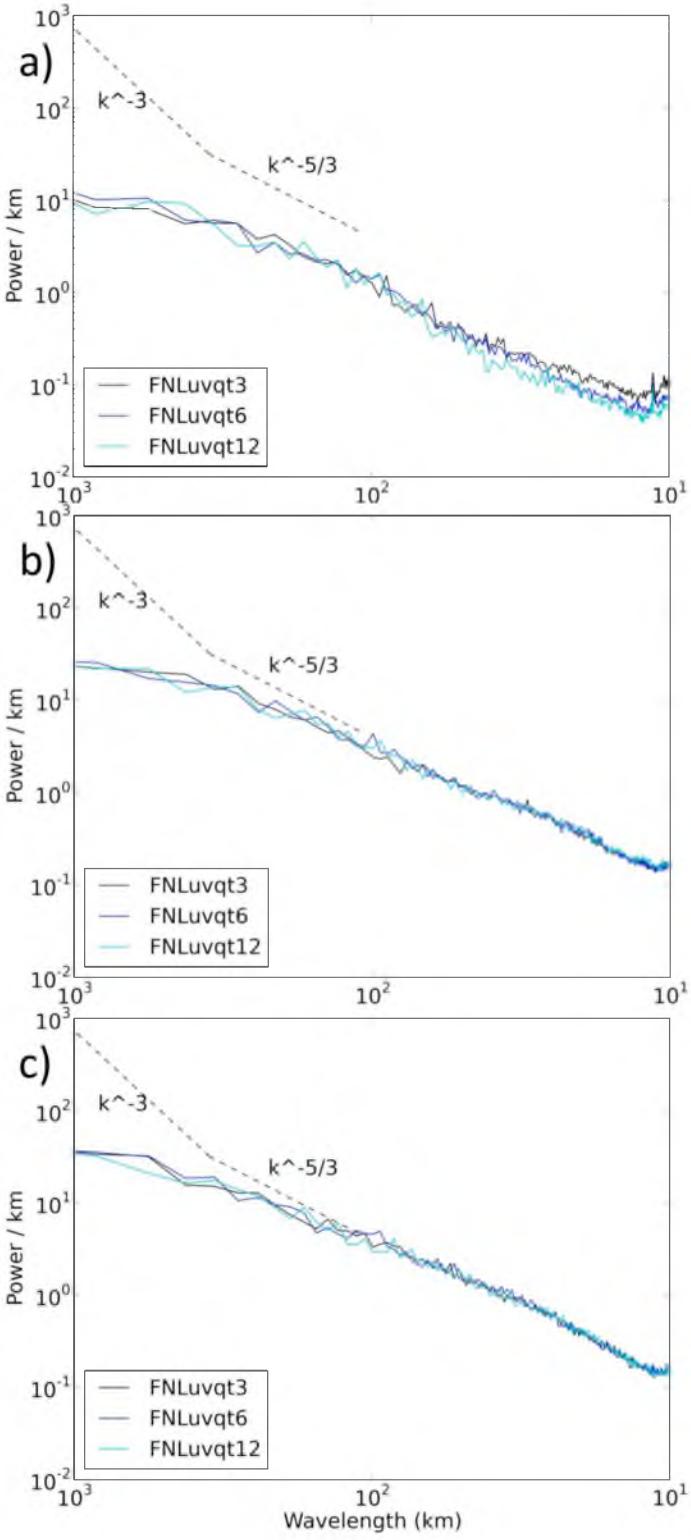

Figure 38. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height perturbation from the ensemble mean, then across a zonal and ensemble average, for FNLuvqt3, FNLuvqt6, and FNLuvqt12 at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

see that it is greatest at wavelengths greater than 150 km and below roughly 70 km. This underdispersion is highest at 0300 UTC, when the convexity and concavity are at their peak, and then decreases as both error and perturbation spectral lines become more linear. The decrease in underdispersion slows and stops as error and perturbation spectral structures saturate near 0900 to 1200 UTC (not shown).

Despite the similarity among these FNLuvqt ensembles, at 0300 UTC, there is slightly higher energy in the FNLuvqt3 perturbation spectra at small wavelengths compared to the ensembles with 6 and 12 h breeding cycles. Noting that the UTC the FNLuvqt6 and FNLuvqt12 perturbation power at 20 km wavelength does not catch that of FNLuvqt3 until 0600 UTC, it appears that the latter was quicker at establishing small-scale perturbation structures, which is what would be expected considering the time scale of baroclinic versus convective style error growth and the quick FNLuvqt3 cycle period. Overall, compared to the atmosphere's natural mesoscale spectral slopes, and $k^{-5/3}$, it appears that over much of the spectrum, the perturbation energy is indicative of faster than $k^{-5/3}$ growth at wavelengths greater than 100 km, and perturbation growth similar to the convective-style slope at wavelengths less than 100 km. While there is much information in Figures 37 and 38, one of the main upshots is the fact that breeding with varying cycling periods does not change the spectral structures of the errors or perturbations significantly. Nevertheless, at 0300 UTC, the FNLuvqt12 ensemble does have notably lower error energy and the FNLuvqt3 ensemble has notably higher perturbation energy, at wavelengths below 100 km, compared to the other FNLuvqt ensembles.

The implementation of a simple vertical localization

As has been mentioned, to accomplish our error breeding in this study we typically use a spatially-invariant scaling factor based on the domain mean of the variable being bred. This follows the strategy used by McLay et al. (2008), wherein he achieved results comparable, in terms of error, to the more sophisticated ET scheme for forecast periods less than 48 hours. In order to test possible effects of this domain and height-invariant scaling factor, we create a height localization method embedded in our breeding scheme. Here we choose to localize vertically rather than in terms of geography. This is done because our simulations are done over a relatively small meridional extent, such that the error growth changes between the barotropic atmosphere near the equator and the baroclinic midlatitudes are not an issue. The height localization is done such that, when scaling back the bred errors for the variable of interest, we use a domain mean which also averages the values over specific height levels. See the top of Table 7. For example, we now calculate separate domain averages for the 800-1000 hPa, 800-500 hPa, 500-200 hPa levels, and another average for all model levels above 200 hPa. Considering the large differences in expected vapor error growth between these levels, it is expected that this vertical localization will provide notably different vapor fields, compared to the non-localized breeding, at the end of the breeding period. Because of this important change in vapor with height, in the vapor localization ensemble, we will breed water vapor and, as is custom in the study, set the breeding cycle to the default 3 h. For ease in discussion, we will call this the FNLq3_height ensemble and the nonlocalized ensemble of comparison is FNLq3 (Table 2).

Effect of vertical localization on water vapor error growth

We now examine the relative error growth, by averaging layer, throughout the 24 h breeding period. The results for member 1 only are seen in Table 7; this simplification is used because of the fact that the rescaling values are quite similar across the members. For simplicity, here we only show rescaling values in the bottom three averaging layers. The corresponding scaling values for the nonlocalized FNLq3 ensemble are also presented for comparison. First, recall that faster error growth means a *smaller* rescaling factor. What we see in the data, as expected, is the fact that error growth in the bottom 200 hPa of the atmosphere is notably higher than it is in the middle atmosphere. Over the first 3 h of the simulation, domain-averaged vapor error growth is 54% (1-1/.65), of the initial perturbation from FNL, for the 800-1000 hPa layer and only 20% (1-1/.83) for the vapor breeding that does not use a localization. Notice how at the start of the breeding period the vapor values in member 1, in the layers above 500 hPa are actually contracting towards the analysis, which is a fairly common issue with ensemble design, and has to do with the perturbations projecting onto contracting directions in the model evolution phase space (J. Anderson, personal communication 2013). From 0000 to 0300 UTC, this occurs for each of the 500-800, 200-500, and < 200 hPa layers (the latter is not shown). Over time, this contraction stops, first in the lower layers, and then higher up, such that by the end of the period member 1, from 200-500 hPa is finally growing away from the analysis in terms of vapor, albeit only by 5% over the previous 3 h. This is compared to a 67% error growth at the same time from 800-1000 hPa, 20% growth in the 500-800 hPa layer, and 30% error growth overall in the nonlocalized ensemble. Thus, the large differences in vapor error scaling factors among the vertical levels implies that a localization factor will be able to account for much more of the particular error idiosyncrasies in each of the four levels.

Table 7

Comparison of scaling factors for member 1 of

FNLq3 and FNLq3_height ensembles during 24 h breeding period

Time	FNLq3_height			FNLq3
	800-1000 hPa	500-800 hPa	200-500 hPa	
03Z	0.65	1.03	1.13	0.83
06Z	0.59	0.88	1.02	0.75
09Z	0.59	0.91	1.07	0.77
12Z	0.60	0.84	1.00	0.76
15Z	0.55	0.83	.97	0.73
18Z	0.49	0.79	.90	0.66
21Z	0.56	0.85	.96	0.73
00Z	0.60	0.83	.95	0.77

In terms of the 3 h error growth over time, in the 24 h period, it appears that these rates plateau around 1800 UTC for each of the vertical levels in the FNLq3_height ensemble and also in the nonlocalized FNLq3 ensemble. After this time, error growth rates trend down slowly in each level and in both ensembles of interest.

Initial condition differences due to vertical localization

Plotted in Figure 39 is the vapor field bias, from the analysis, averaged between 800-1000 hPa at the end of the breeding period for the FNLq3 ensemble (0000 UTC 24th). What we see is a notable warm bias in the northwest portion of the domain for almost all members. Differences, however, arise in the spatial patterns of this area of warm bias and other features across the domain. For comparison, a similar 800-1000 hPa vapor bias calculation, but the localized FNLq3_height ensemble is presented in Figure 40. Note that it appears the localization is performing correctly, as the low-level bias of the localized, FNLq3_height, ensemble is lower than that of the FNLq3 ensemble, which uses the invariant rescaling factor. For example, in FNLq3, the bias values in the northwest corner of the domain often reach 1-1.5 g/kg, whereas in the FNLq3_height, ensemble bias values in the same region only reach 0.5-1 g/kg. Note how the spatial patterns are largely the same despite the large differences in perturbation magnitude. Of course, we expect the differences between the FNLq3 and FNLq3_height ensembles to change significantly at higher levels.

Figure 41 shows the vapor bias from 500-800 hPa for the FNLq3 ensemble and Figure 42 shows the same for the FNLq3_height ensemble. First, in Figure 41, we see the large scale moist perturbation has shifted to the center of the domain, with most members showing a 2-3 g/kg moist bias there at 0000 UTC 24th Aug. When using a localization

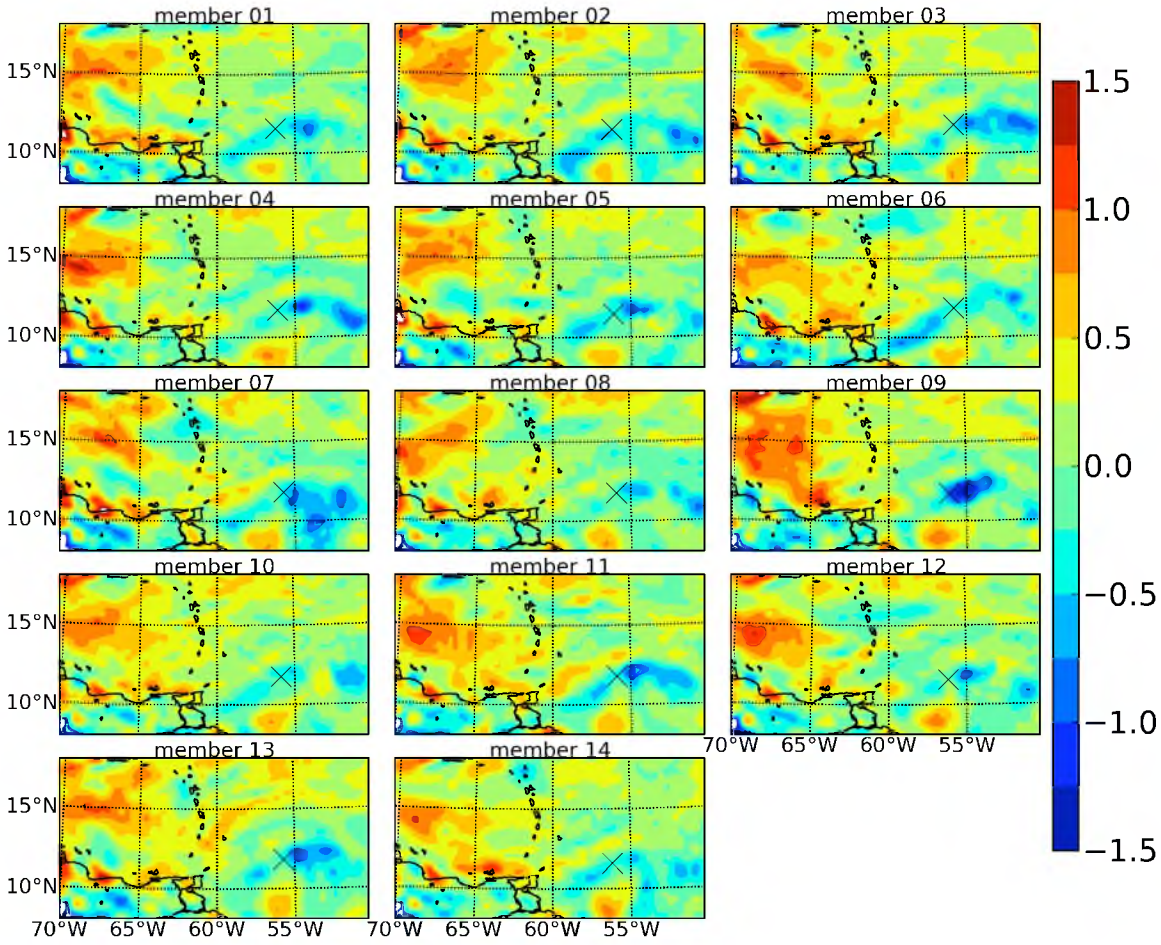

Figure 39. Bias of 800-1000 hPa vapor (g/kg) for each member of the FNLq3 ensemble (nonlocalized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

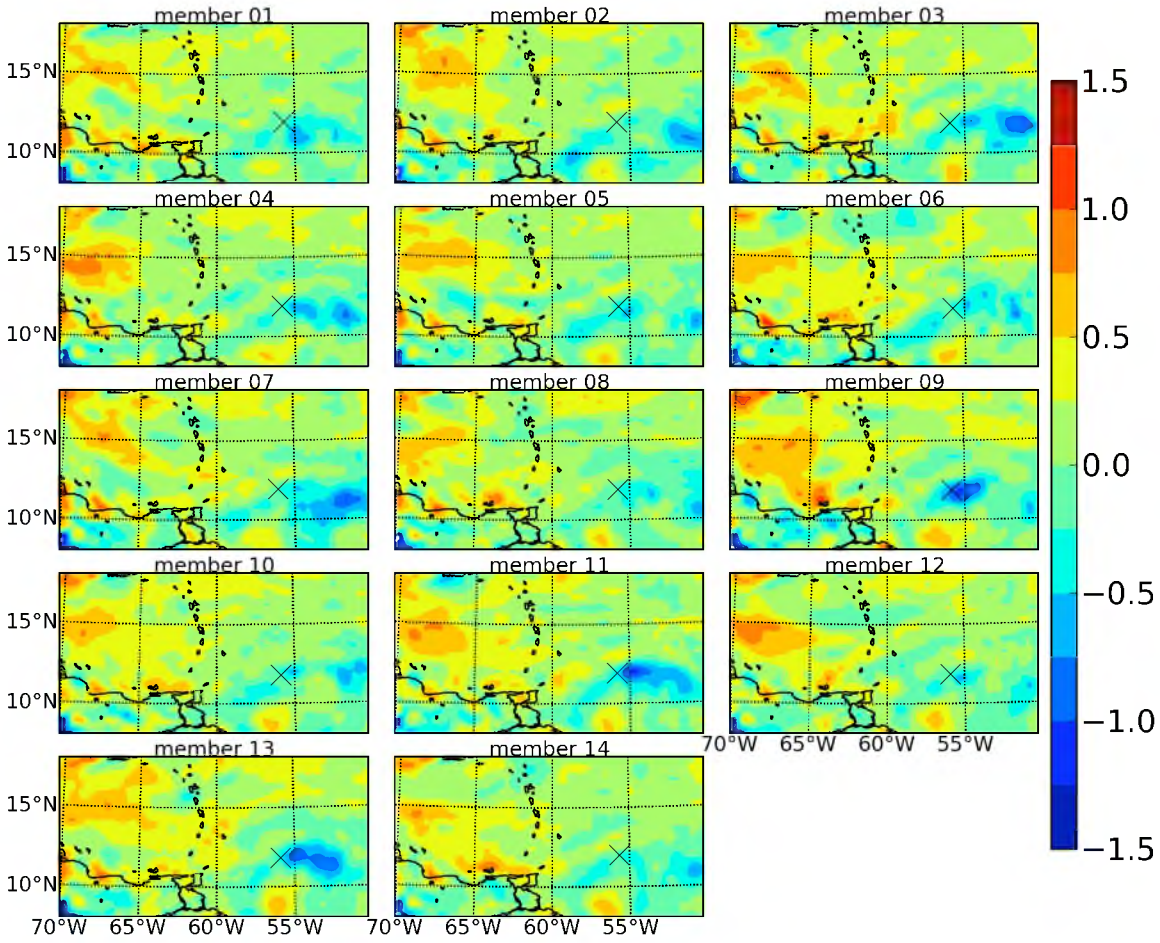

Figure 40. Bias of 800-1000 hPa vapor (g/kg) for each member of the FNLq3_height ensemble (localized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

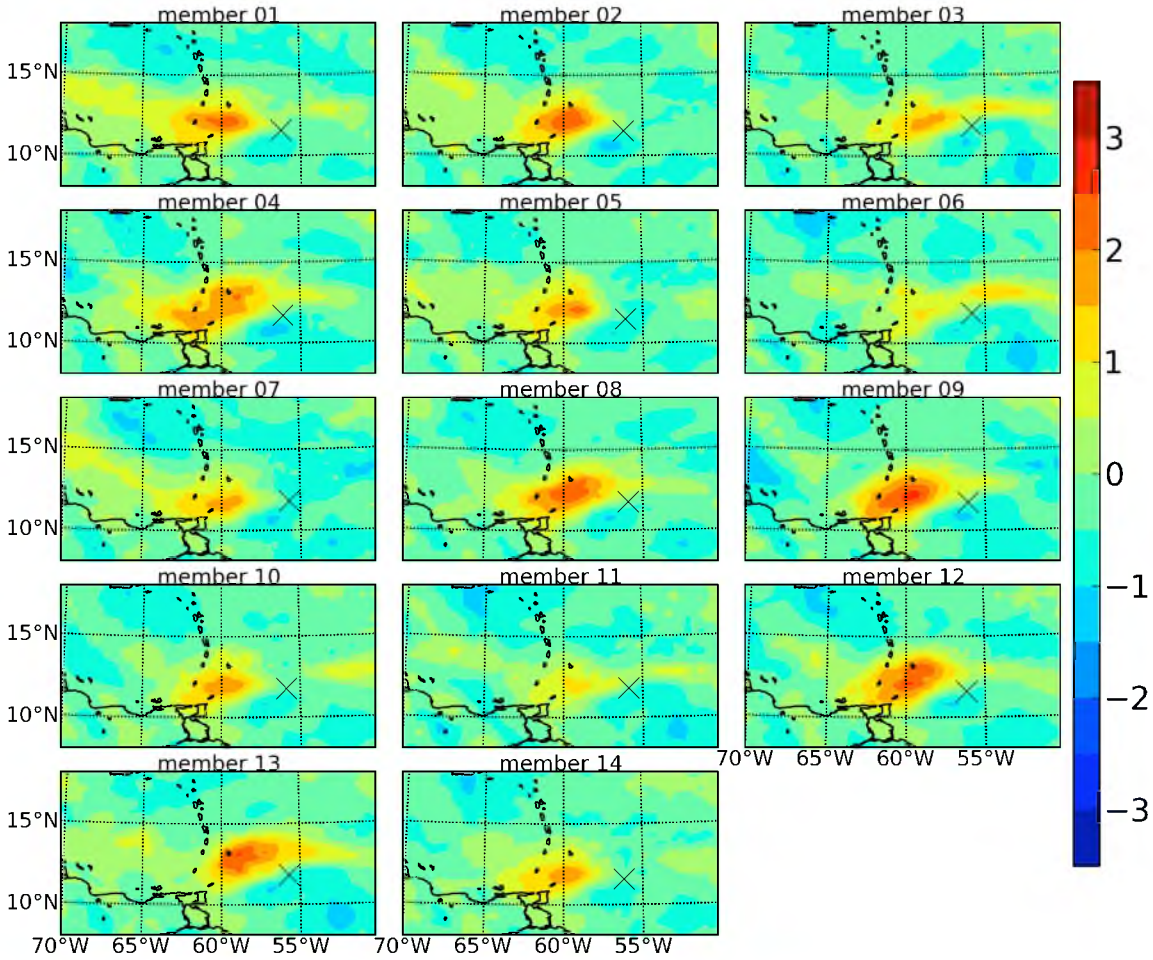

Figure 41. Bias of 500-800 hPa vapor (g/kg) for each member of the FNLq3 ensemble (nonlocalized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

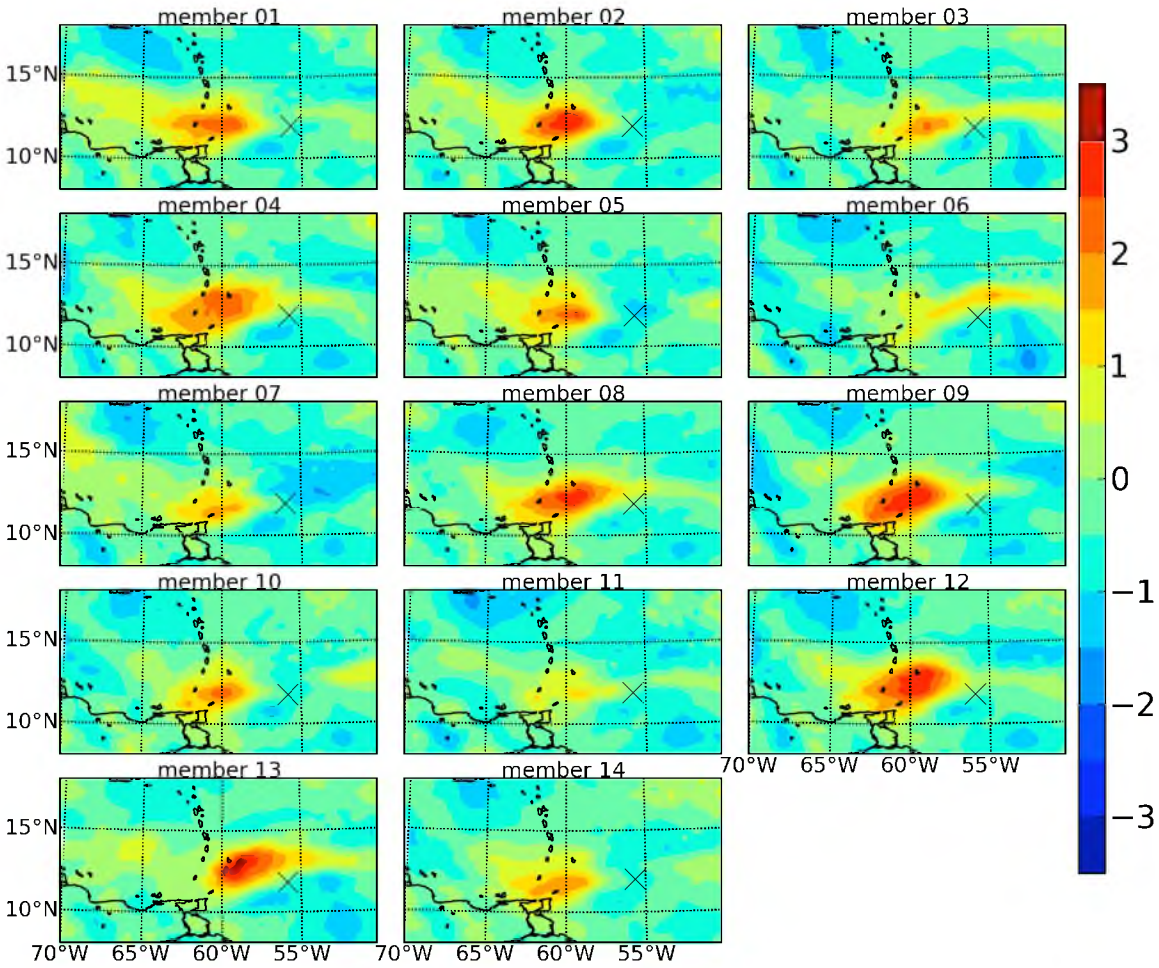

Figure 42. Bias of 500-800 hPa vapor (g/kg) for each member of the FNLq3_height ensemble (localized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

(Figure 42), we see that the magnitudes of the perturbations are actually slightly larger than they are in the FNLq3 ensemble. Again, this is sensible, as when using a spatially-*invariant* rescaling factor, perturbations at certain height levels would be rescaled by too little. While there is a consistent moist perturbation near the center of the domain, toward the southeast, near the location of the nascent disturbances, there is actually a consistent 500-800 hPa dry bias in the FNLq3_height versus the normal FNLq3 ensemble. It is suspected that this lack of moisture may play a role in the differences in genesis likelihood between the two ensembles and this will be further detailed below. While the perturbation magnitudes from 500-800 hPa are larger in FNLq3_height versus FNLq3, the spatial structure of the localized ensemble bias, from member to member, remains largely the same as the nonlocalized ensemble. This is interesting because one might imagine the scale of the perturbations over certain parts of the domain, early in the breeding period, as having an effect on perturbation patterns at the end of the breeding period. Largely, however, this does not occur.

Member bias for the 200-500 hPa level in the FNLq3 and FNLq3_height ensembles is plotted in Figures 43 and 44. Perturbation spatial patterns for both these ensembles at this height are quite similar to those from 500-800 hPa, in that there is a notable warm bias, for almost every member, near the center of the domain. Comparing Figure 43 with 44, we see that the localization at this height pushes the members notably further from the analysis across much of the domain, compared with what occurs lower in the model. Despite the height levels of interest, vapor perturbations in the localized FNLq3_height ensemble often reach 1-1.5 g/kg over parts of the domain. Overall, while the relative contraction and expansion of perturbations occurred as expected over the various heights, the different breeding styles produce more similar spatial patterns for the same member, between the ensembles, than expected. In terms of the FNLq3_height vapor bias over

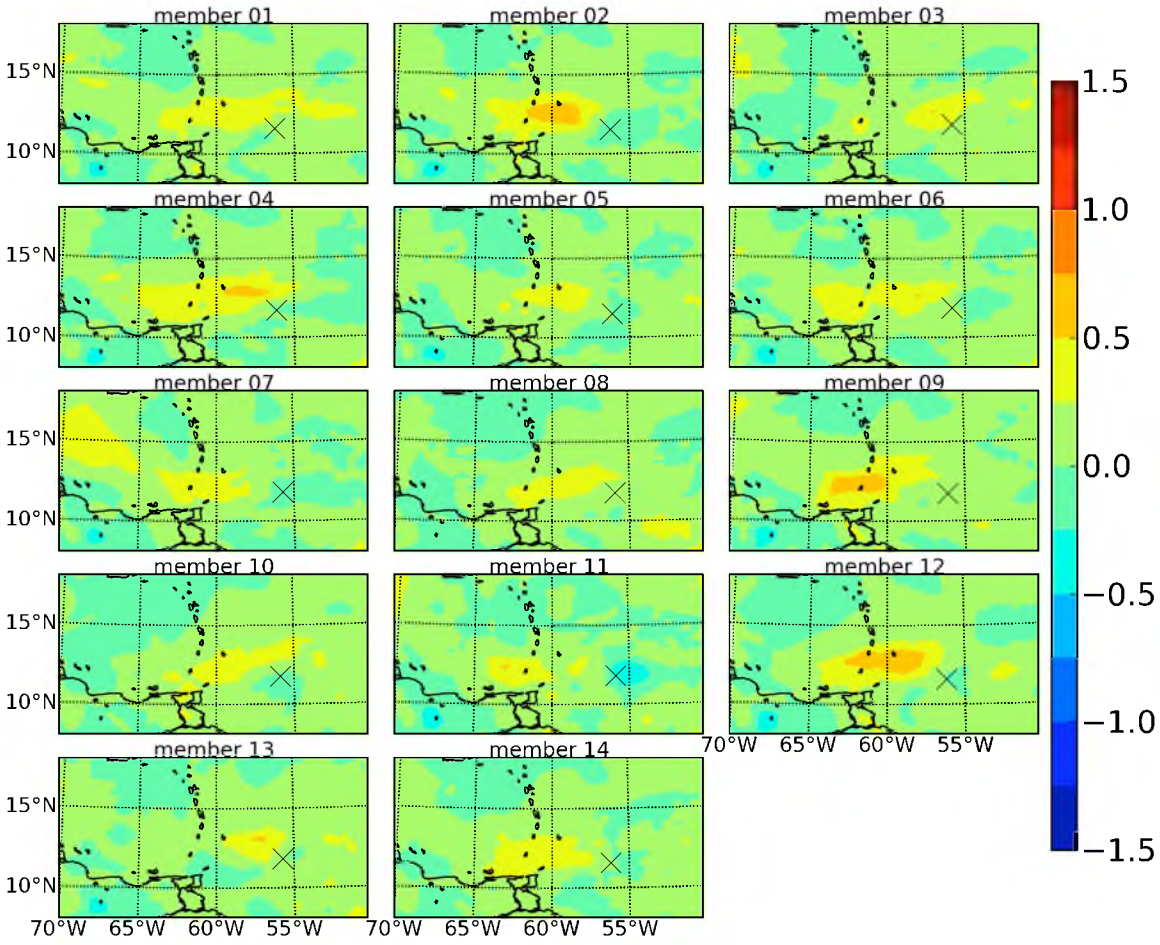

Figure 43. Bias of 200-500 hPa vapor (g/kg) for each member of the FNLq3 ensemble (nonlocalized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

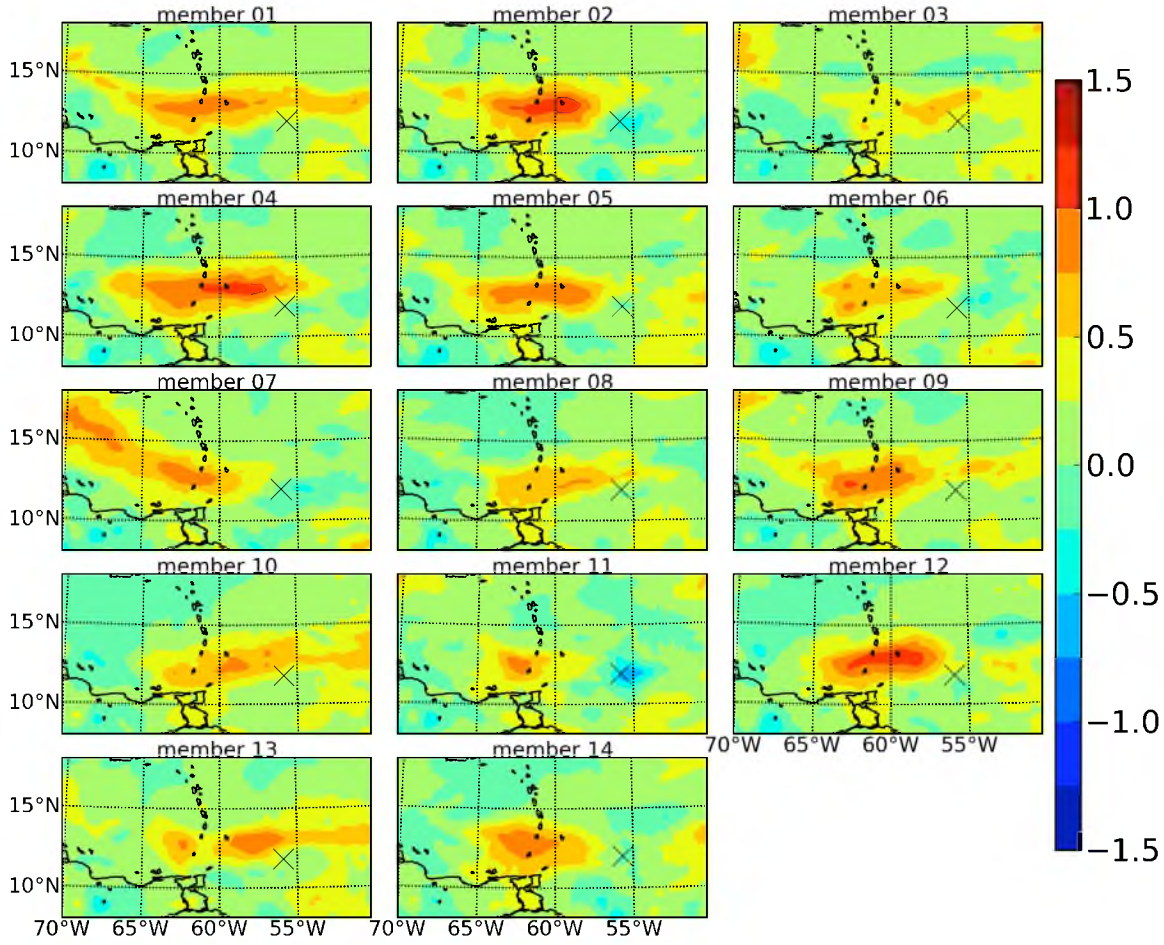

Figure 44. Bias of 200-500 hPa vapor (g/kg) for each member of the FNLq3_height ensemble (localized) at 0000 UTC 24th (end of breeding period). X marks the location of the pre-Ernesto disturbance in each member.

time, we find that it is slightly better than FNLq3 at 500 hPa (Figure 7), and worse or roughly equal to FNLq3 at 925, 850, and 700 hPa (Figure 4-6). In general, these are not the kind of improvements that one would expect from a vertical localization, which adds credence to the fact that invariant scaling factors can be used quite effectively.

Synoptic and statistical comparisons of localization-related genesis

Moving toward the impact of the different rescaling factors, and thus initial perturbations, on TC genesis, for each member of the FNLq3_height ensemble a synoptic view of 850 hPa vorticity and geopotential height at the actual genesis time is presented in Figure 45. This is comparable with Figure 10, where the same calculations were made for the FNLq3 ensemble. In general we see that the nonlocalized ensemble produces more established TCs than FNLq3_height by 1800 UTC 24th. This is displayed in Table 8, where we see that by this time, the nonlocalized ensemble has produced two TCs above the general thresholds, while the localized, or FNLq3_height ensemble shows no TCs having formed by 1800 UTC 24th. As alluded to, this reduced likelihood of genesis predicted by the vertical localization ensemble, small though it is, may be explained by the reduced amount of water vapor at the initial time (Figure 42), in the critical 500-800 hPa layer near the disturbance around 11°N and 56°W, as compared to the FNLq3 ensemble (Figure 41). This reduced amount of near-disturbance water vapor appears to only affect initial TC organization, as 18 h later at 1200 UTC 25th August (Table 8), all members of both ensembles exhibit genesis. Not only have the TCs in each ensemble achieved genesis by that time, but the intensity of the storms between the FNLq3 and FNLq3_height ensembles are quite similar. Note that while the TC genesis difference by 1800 UTC 24th is only two, at this point the intensity differences from one ensemble to the other are fairly consistent. These impacts of environmental precursors on TC genesis

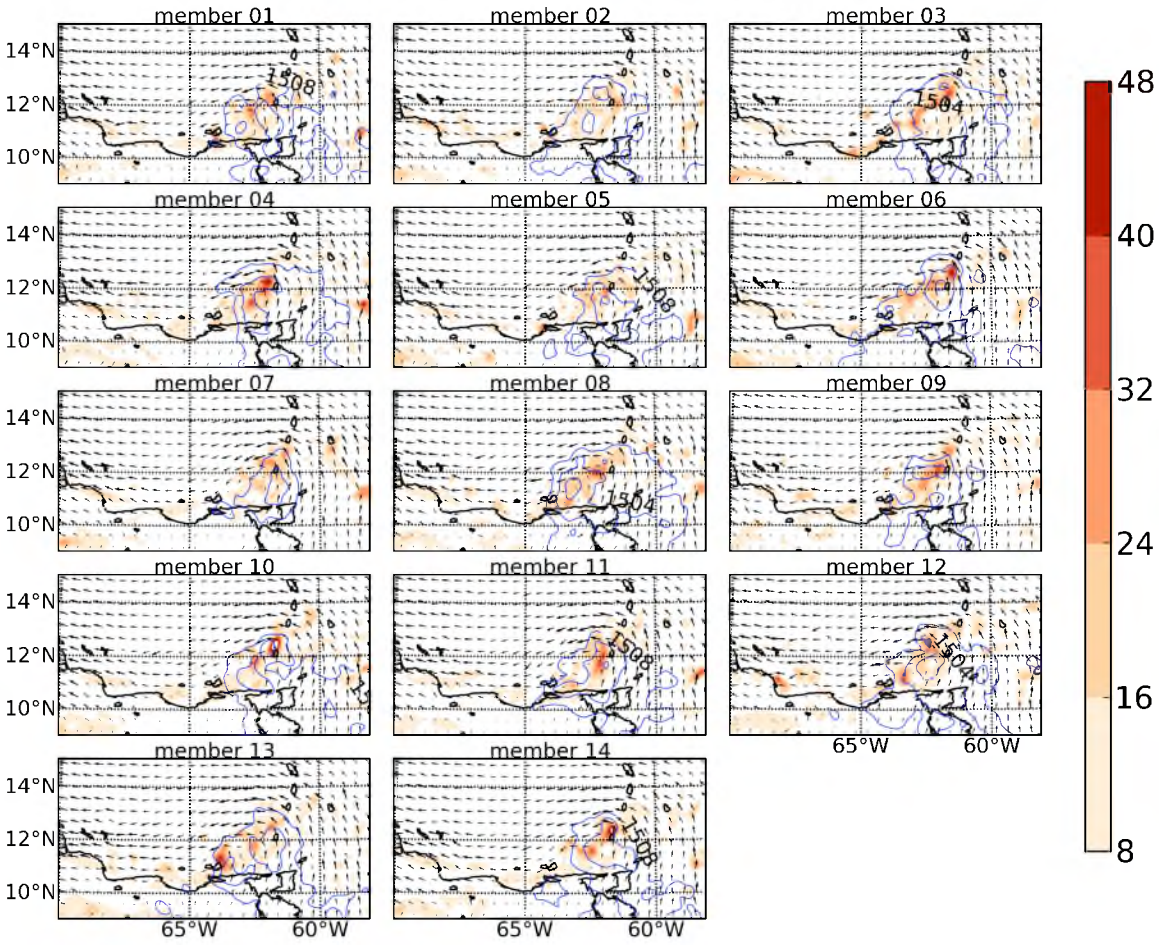

Figure 45. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLq3_height ensemble members at 850 hPa at 1800 UTC 24th.

Table 8

Genesis statistics for

Hurricane Ernesto (forecast began 0000 UTC 24th)

Ensemble experiments	18Z 24th	12Z 25th
FNLq3	02/14	14/14
FNLq3_height	00/14	14/14

will be further examined in chapter 7.

Ensemble spread and dimensionality in localization ensemble

In terms of the near-TC error dimensionality throughout the ensemble period for the localized and nonlocalized FNLq3 ensembles, the results are presented in Figure 46. We see that both of the bred ensembles have a slightly higher vapor E-dimension by 9 h into the simulation period compared to the control ensemble. Also compared to the control ensemble, the bred ensembles show a more notable decline in dimensionality from 12-24 h into the simulation period. These bred and control ensemble differences have been noted before. What is new in this figure is the fact that, despite the significant differences in initial perturbation size, the FNLq3 and FNLq3_height ensembles show very similar E-dimension near the TC and throughout the simulation period. Not only are these two ensembles similar in terms of dimensionality but also in terms of near-TC vapor spread. While the similar spatial perturbation structures between the localized and nonlocalized ensembles would not be expected to produce notably different forecast error dimensionality, 850 hPa ensemble spread on the other hand would be expected to change due to the notably different perturbation magnitudes of each ensemble. This is the case even though 850 hPa appears to be where the localization effect on perturbation size changed from negative (below 850 hPa) to positive (above 850 hPa). Having said this, it is notable how small of change in vapor spread occurs at this level over time, especially since we are only examining near-TC values here in the FNLq3 and FNLq3_height ensembles.

Because of this lack of change in either vapor dimensionality or spread when using

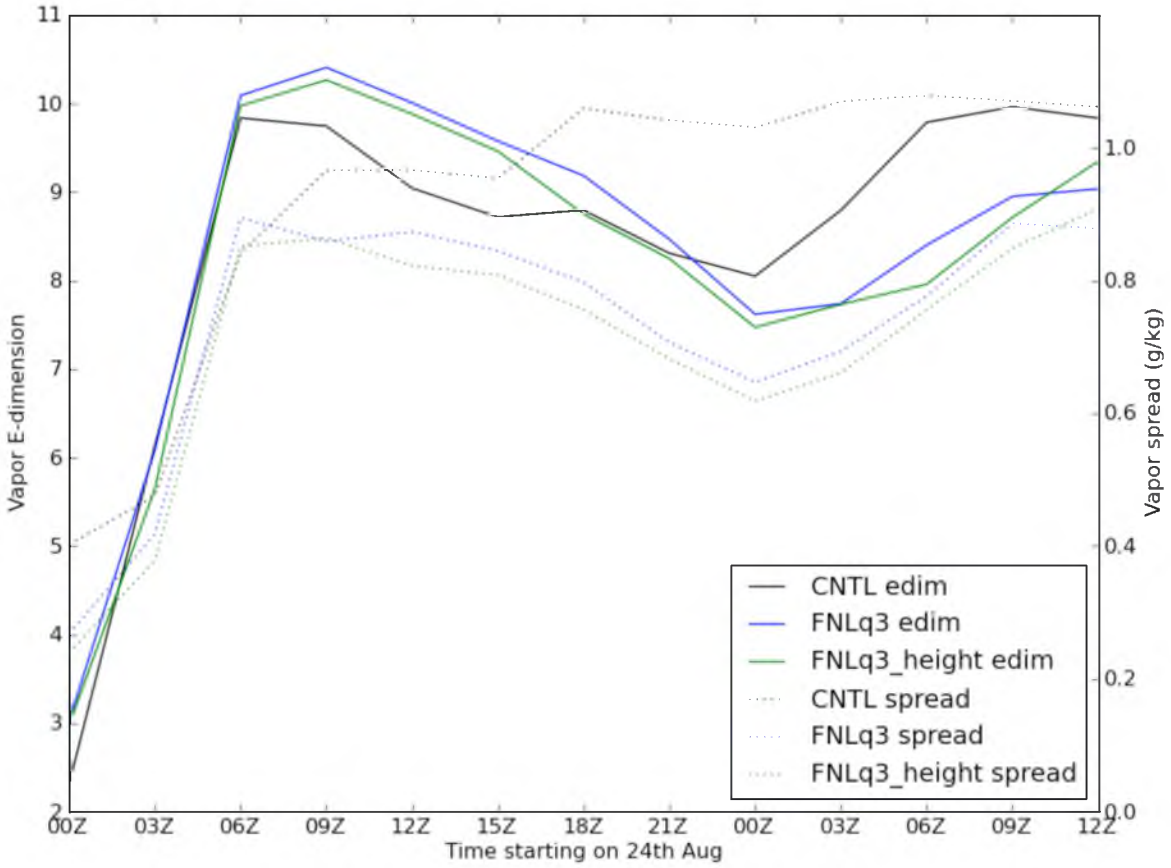

Figure 46. Time series of error E-dimension and spread (g/kg) for 850 hPa water vapor within 200 km radius of simulated TC center in the CNTL, FNLq3, and FNLq3_height ensembles.

a vertical localization, we also would not expect much of a change in the correlation between these two metrics over time. We perform a 1000-sample bootstrap of this correlation for both ensembles and plot the cumulative distribution in Figure 47a, with that of the control ensemble for comparison. Note how these particular breeding schemes do provide a much stronger relationship between dimensionality and spread than does the control ensemble. Also, note the close similarity in bootstrapped-correlation values between the FNLq3 and FNLq3_height ensembles. Plotted in Figure 47b is the 1000-sample bootstrapped correlation between *domain-averaged* ensemble vapor root mean squared error (RMSE) and spread for the FNLq3 and FNLq3_height ensembles at 850 hPa. When constructing an ensemble, it is very difficult to get it to properly align error with spread (Berner et al. 2009). This is seen in Figure 47b, as there are many bootstrapped samples where the correlation between RMSE and spread for these ensembles is below or near zero. When comparing the ensembles of interest, we see that the localization factor appears to slightly, but generally, increase the correlation between domain-averaged 850 hPa error and spread. Overall, while the vertical localization factor 1) did not alter ensemble dimensionality because it did not alter the spatial patterns of the error and 2) made it more difficult to spin up a TC probably because it reduced vapor in key regions, it did manage to increase the correlation between ensemble spread and RMSE. It is also found (not shown) that the spectral spatial patterns for the FNLq3 and FNLq3_height ensembles were almost identical.

Discussion

We find that doubling the breeding cycling period from 3 to 6 h increased error growth by only 0-50%, thus demonstrating that halving one's number of cycles does not result in a corresponding increase in error growth per cycling period. Thus, the growth-

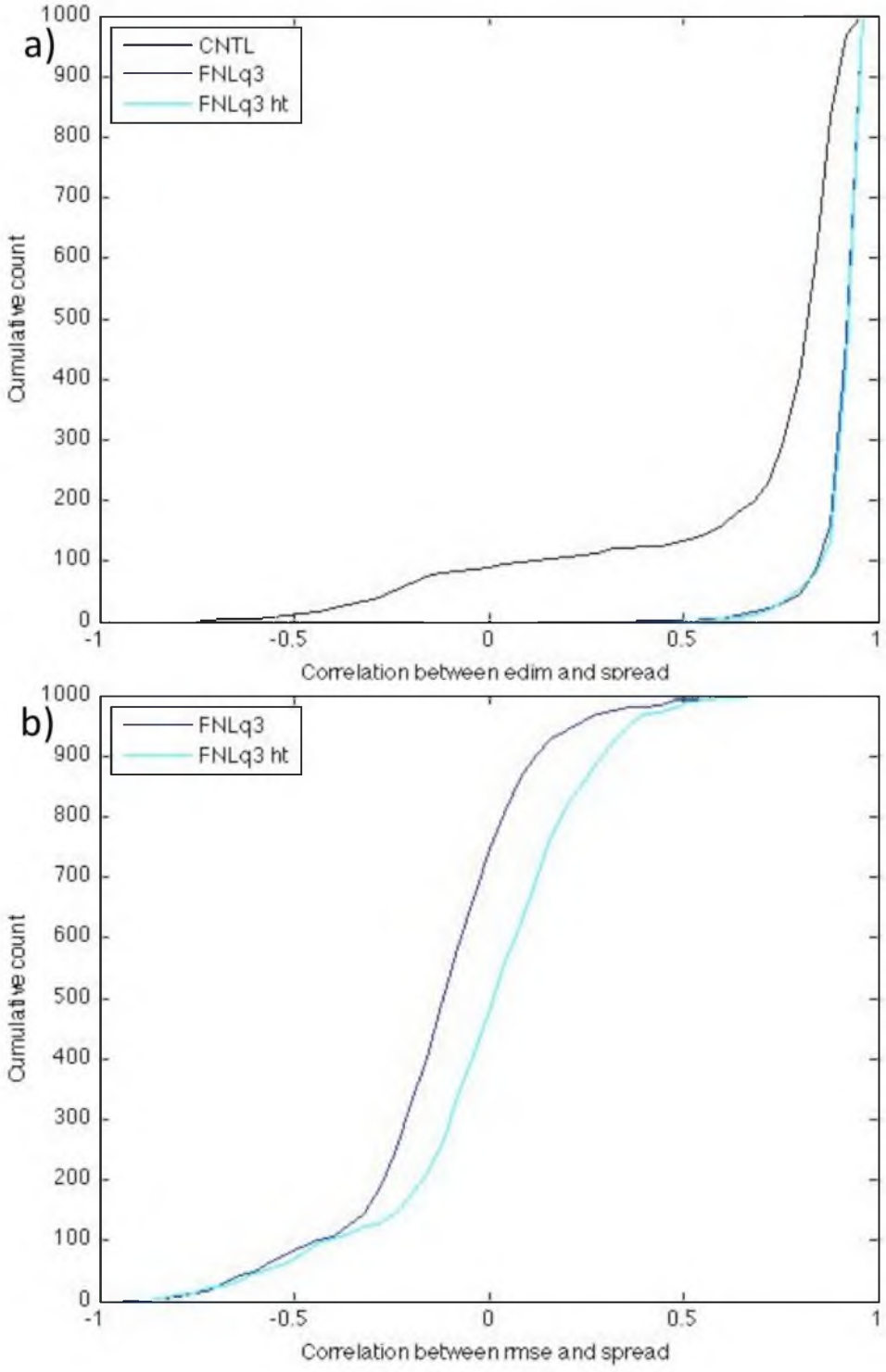

Figure 47. Cumulative distribution of bootstrapped correlation coefficients between FNLq3 and FNLq3_height ensembles for 850 hPa water vapor between a) E-dimension and spread within 200 km radius of simulated TC center and b) domain-averaged root mean squared error and spread over time. Correlations have been found using a 1000 sample bootstrap.

per-cycle vs number-of-cycles tradeoff is more conducive to the goals of the breeding scheme at 3 h cycles compared to 6 or 12 h cycling periods. This, of course, is based on forecasts made in the Tropics, centered around a tropical cyclone. Midlatitude-based simulations may result in different conclusions.

The correlation between spread and E-dimension near the TC was enhanced by longer breeding periods compared to short, and decreases when breeding multiple variables instead of just one. While we do not fully understand the implications of this enhanced dimensionality and spread relationship, to our knowledge, this is the first time it has been examined in the context of varying error growth regimes in the Tropics. Overall, these ensemble dimensionality changes do have important implications for determining how easily a particular forecast can be improved by enhancements along a particular variance direction (Bishop and Toth 1999).

A vertical rescaling localization was employed in this project and it did successfully create smaller perturbations near the surface and larger perturbations in the midlevels. The localization appears to reduce the likelihood of TC genesis because of the fact that it reduced water vapor levels near the TCs in the critical 500-800 hPa layer. Further, this technique did not notably reduce model bias for either vapor or θ at any height from 500 hPa to the surface, as compared with a non-vertically-localized ensemble. This finding leads us to believe that for this case, the invariant scaling factor as used by McLay et al. (2008) provides a good tradeoff in terms of accuracy to computational cost.

Highlights of Chapter 4

- Doubling the breeding cycling period from 3 to 6 h only led to error magnitudes which were 0-50% greater. Thus, halving one's number of cycles does not result in a corresponding increase in error growth per cycling period. Because of this, the growth-per-cycle versus number-of-cycles tradeoff is more conducive to efficiently sampling the underlying forecast error growth with 3 h cycles compared to 6 or 12 h cycling periods.
- Breeding multiple variables reduces the significant error dimensionality decrease over time compared with FNLq3 and FNLth3 ensembles.
- In addition, breeding multiple variables instead of just one notably weakens the relationship between the ensemble's spread and dimensionality.
- Vertical localization did not notably reduce bias in lower levels.
- Reduced water vapor from 500-800 hPa near the initial disturbance appears to reduce likelihood of TC formation.

CHAPTER 5

THE IMPACT OF BOUNDARY CONDITIONS AND SIZE OF INITIAL PERTURBATIONS

The impact of boundary conditions

Throughout these experiments, the boundary conditions have been exclusively FNL, with only the initial conditions changing with each ensemble. This has allowed us to focus on the differences in initial conditions. To test the impact of varying boundary conditions, GEFS boundary conditions, one distinct set for each of the members, are now used in an ensemble forecast and we thus create a FNLuvqt3 ensemble with such BCs (called FNLuvqt3_GEFSBCs; Table 2) and compare it against the FNLuvqt3 ensemble with the standard FNL BCs.

First, we examine the comparative bias of the FNLuvqt3 versus FNLuvqt3_ GEFSBC ensemble (Figure 8) and find that the latter ensemble shows a slightly higher bias for vapor from 400-700 hPa and significantly higher bias for θ above 800 hPa; elsewhere, the differences are mostly negligible. Overall, however, our choice of FNL BCs for the majority of our simulations appears prudent. Next, we evaluate the accompanying differences in terms of synoptic fields 18 h into the simulation period. To do this, we plot 850 hPa vorticity, geopotential height, and wind vectors for this FNLuvqt3_GEFSBCs ensemble in Figure 48. This is comparable to the corresponding FNLuvqt3 ensemble in Figure 30. What we see is a similar TC positioning

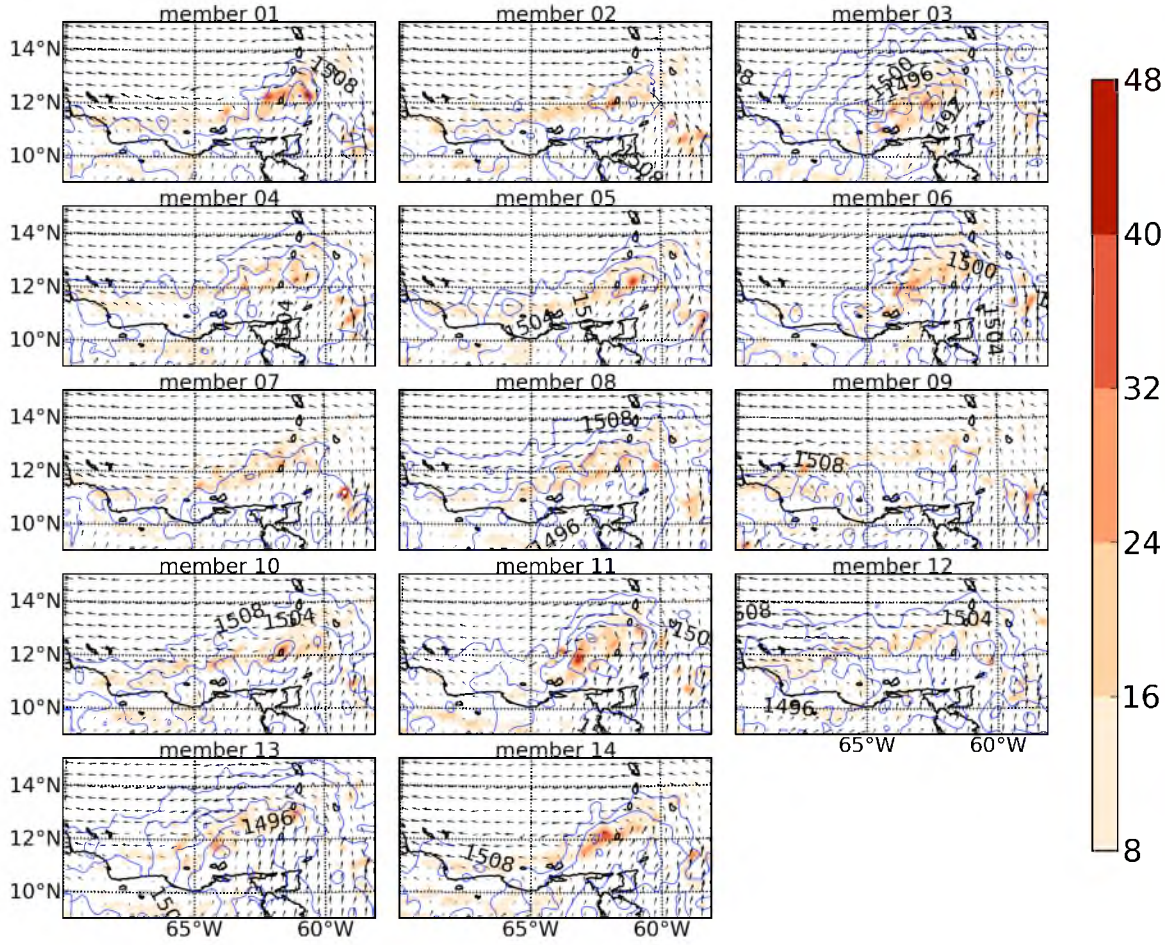

Figure 48. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt3_GEFSBCs ensemble members at 850 hPa at 1800 UTC 24th.

and intensity spread between the members of both ensembles at this time. To better compare average ensemble intensity, we take the ensemble mean for each and plot the results in Figure 49. It is quickly obvious that the ensemble with GEFS boundary conditions produces slightly stronger TCs overall. If one looks closely, this can be seen most notably in terms of geopotential height differences between Figure 30 and 48. Despite the stronger TCs produced by the GEFS BCs, however, this ensemble still exhibits a significant amount of diversity among the members, as the disturbances in member 2 and 7 are quite weak, for example (Figure 48).

In terms of spatial spectra of errors and perturbations, the choice of boundary conditions does not have a notable effect (not shown), as the amount of underdispersion is almost identical for both ensembles. Also, in terms of error dimensionality, the two ensembles are essentially identical (not shown). Despite the slightly stronger TCs created by FNLuvqt3_GEFSBC, the notably lower bias of the FNLuvqt3 ensemble confirms that our choice of FNL BCs was appropriate in terms of investigating bias and error growth as manifest in various perturbation methods.

The effect of half size perturbations

In order to determine the effect of perturbation size on the resulting simulations, we create an ensemble of half size perturbations called FNLuvqt3_half, which, for a technical reason, is constructed by performing two extra size rescalings at the end of the normal breeding period. First, in terms of the bias of the half size perturbation ensemble (Figure 8), we find that it is also notably higher than the FNLuvqt3 ensemble, in terms of θ, from 400 to 600 hPa. In terms of vapor, the two ensembles are quite similar, with the bias of the FNLuvqt3_half ensemble slightly higher in the midlevels. Thus, while we

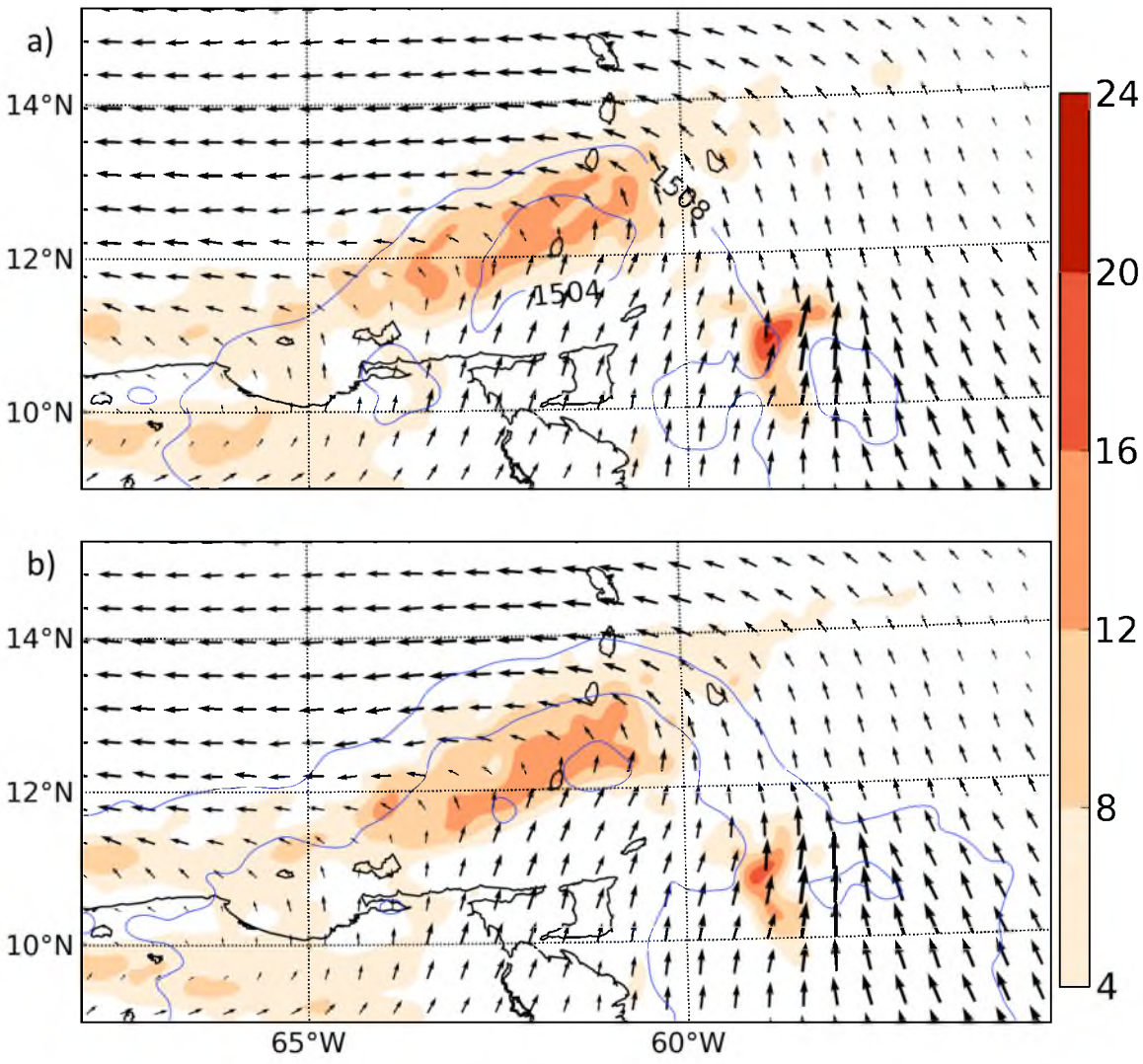

Figure 49. Ensemble averaged vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) at 850 hPa at 1800 UTC 24th for a) FNLuvqt3 with FNL BCs and b) FNLuvqt3 with GEFS BCs.

would expect perturbation size to primarily affect ensemble spread, we see that in our ensemble, it has a clear affect on ensemble bias. The synoptic results of this ensemble are plotted in Figure 50. When comparing this FNLuvqt3_half ensemble with that of the standard FNLuvqt3 ensemble (Figure 30), we find, as expected, that the variability of the former ensemble's TCs' intensity and spatial positioning are dramatically reduced. We find that the reduced perturbations of the FNLuvqt3_half ensemble are not large enough to create meaningful differences between its resultant forecasts and the FNL-based deterministic forecast seen in Figure 51. While we wondered whether, in some situations, starting these ensembles with somewhat unbalanced fields was delaying TC spin-up time, it appears that reducing the perturbation size would hinder the ensemble's dispersion characteristics because it does not provide a fair portrayal of the uncertainty related to this TC's difficult forecast. As a further explanation of the lack of improvement provided by the half size perturbations, when viewing high-resolution (in time) output from one of our bred simulations, we find that over the first hour of forecast time, there are no examples of spurious behavior or unphysical waves in terms of 850 hPa geopotential height at any point across the domain.

To better compare the FNLuvqt3 ensemble with a corresponding ensemble of half size perturbations, we average over the ensemble members and provide a synoptic view of vorticity at TC genesis time (Figure 52). What we find is that the ensemble with half size initial perturbations, compared with the FNLuvqt3 ensemble, shows a TC which is much more highly compact and intense, both in terms of geopotential height and in terms of vorticity. This intensity, of course, has to be weighed against the lack of spread seen in Figure 50. While one of the goals of this study is to produce regional simulations of

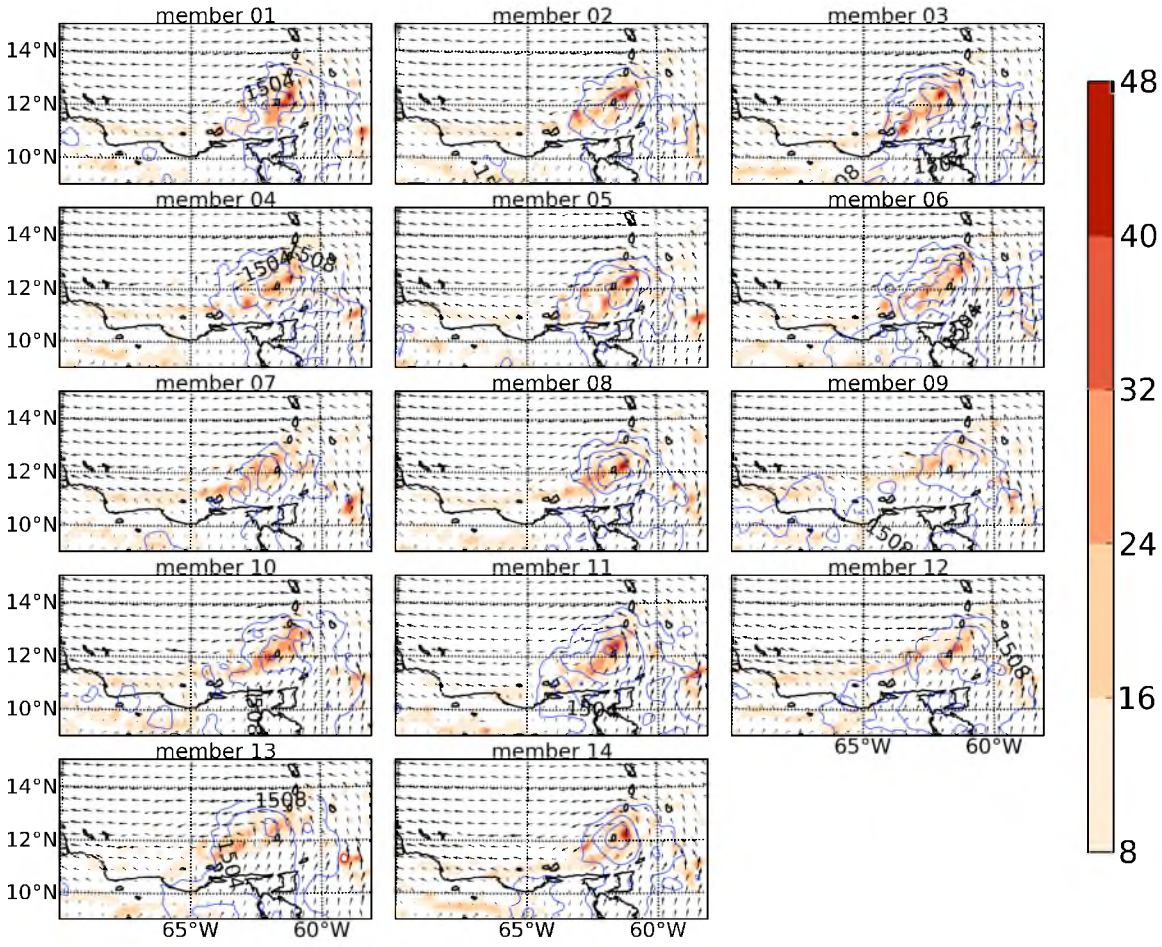

Figure 50. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt3_half ensemble members at 850 hPa at 1800 UTC 24th.

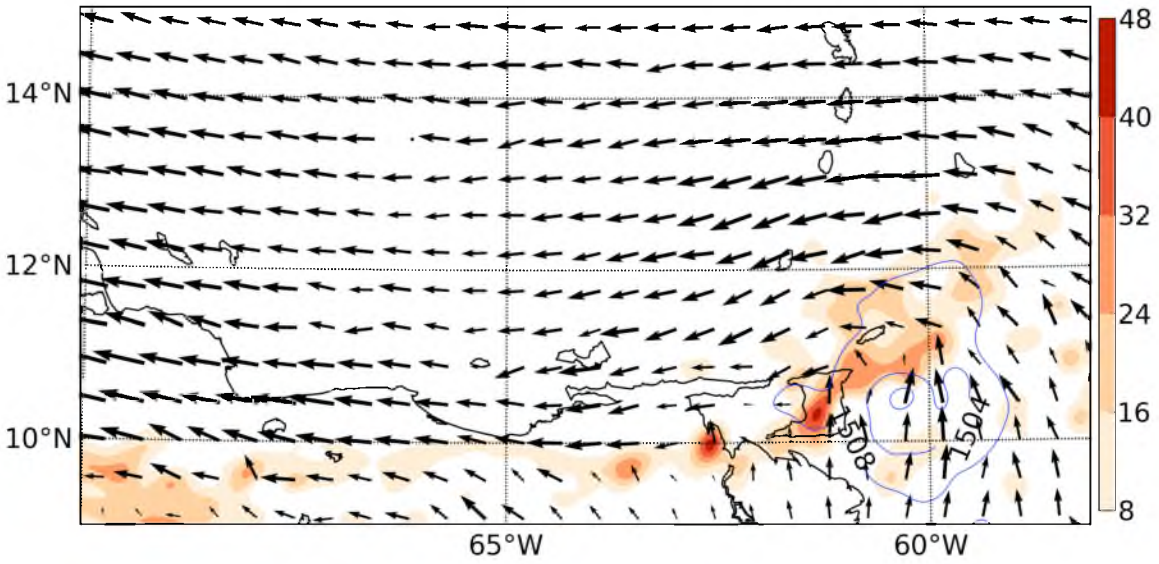

Figure 51. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of Hurricane Ernesto at 850 hPa at 1800 UTC 24th from WRF deterministic forecast using FNL ICs and BCs.

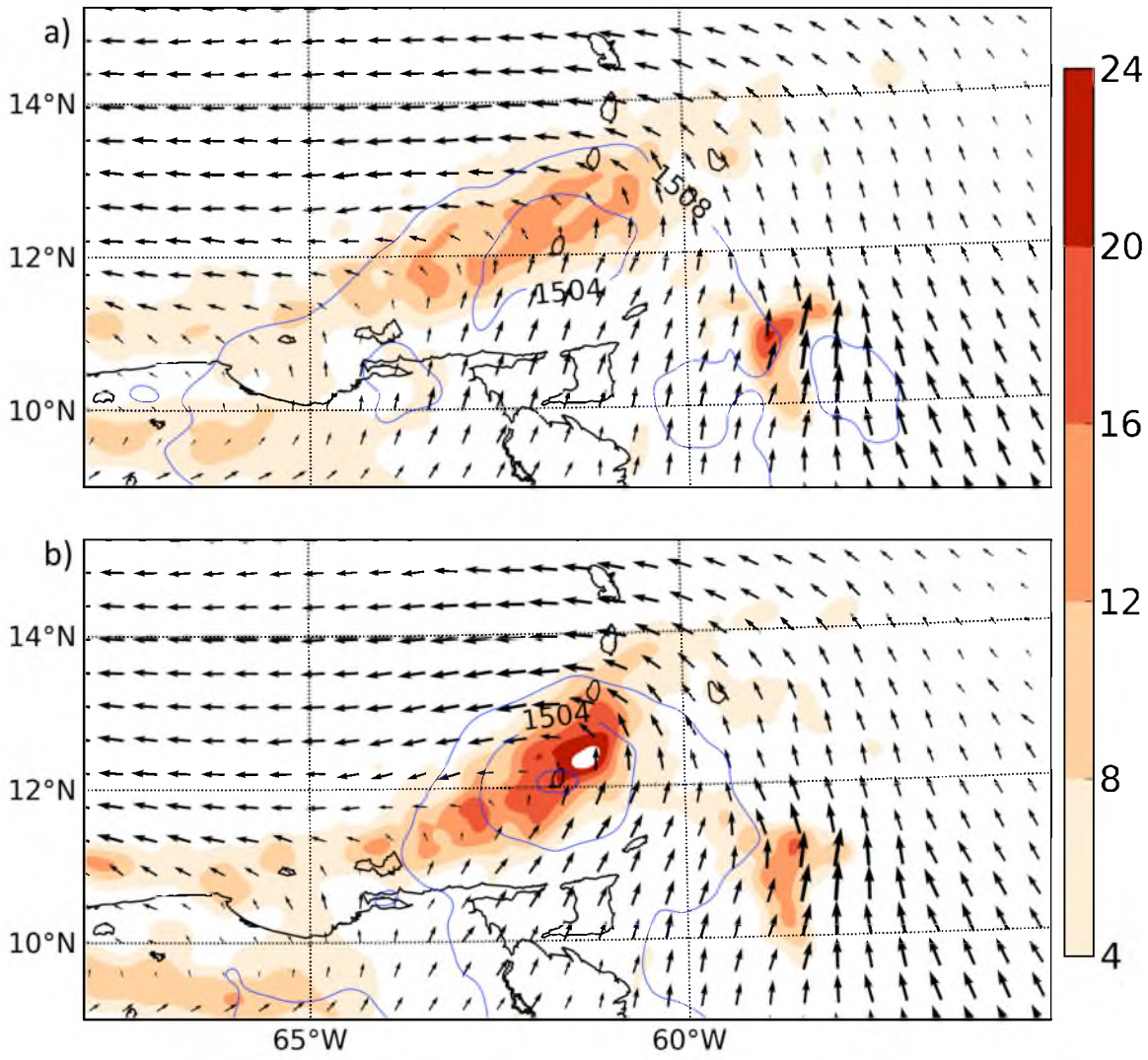

Figure 52. Ensemble averaged vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) at 850 hPa at 1800 UTC 24th for a) FNLuvqt3 and b) FNLuvqt3_half.

Ernesto which forecast genesis and provide spread that matches the associated error, reducing the size of our perturbations does not bring us closer to achieving this goal.

Highlights of Chapter 5

- Boundary condition perturbations have a notable impact on regional ensemble forecasts of TC genesis.
- The FNLuvqt3 ensemble based on FNL BCs produced notably lower ensemble θ bias than did FNLuvqt3 based on GEFS BCs.
- GEFS BCs produced slightly stronger TCs than did a similar forecast of FNLuvqt with FNL BCs.
- The stronger FNLuvqt3_GEFSBCs simulation does not come at the expense of TC intensity spread.
- Half size perturbations increase θ bias from 400 to 600 hPa significantly compared to the FNLuvqt3 ensemble with full size perturbations.
- Using half size perturbations in FNLuvqt does increase forecast TC intensity overall, largely because the members are converging to the deterministic forecast, which produces a strong TC.

CHAPTER 6

A MULTIPHYSICS ENSEMBLE

In the forecasting community, it is common to create an ensemble of model simulations based around the use of different physics parameterizations in order to account for the errors in the individual model physical parameterizations. It is also relatively straightforward to create a myriad of carefully distinguished ensemble members because of the fact that one may isolate the cumulus, microphysics, planetary boundary layer, radiation, and other parameterization schemes quite easily. The multi-physics ensemble we will now discuss represents the beginning of our shift towards ensemble-creation methods that focus on model error instead of initial condition error. For ease of discussion, this ensemble based on varying physics parameterizations will be referred to as PHYS. Its parameterization specifications are noted in Table 9. Note that all of the members use the MM5 similarity surface layer scheme and the 5-layer thermal diffusion land surface scheme.

BGM versus multiphysics: Synoptic and statistical comparisons of genesis

Since the multiphysics ensemble does not have a breeding period nor any unique initial conditions, we start the analysis with a synoptic overview at 1800 UTC 24th August (Figure 53), which is when Ernesto's best track data show genesis. What we see is the fact that there are relatively intense storms in each of the members of the ensemble.

Table 9

Parameterization specifications for PHYS ensemble.

Note that the namelist number for use in WRF is in parentheses

Member	PBL	Cumulus	Microphysics
01	YSU (1)	GD (3)	Lin (2)
02	MYNN3 (5)	GD	Lin
03	ACM3 (7)	GD	Lin
04	UW (9)	GD	Lin
05	MRF (99)	GD	Lin
06	YSU	KF (1)	Lin
07	YSU	BMJ (2)	Lin
08	YSU	Simp. Arak.-Schub. (4)	Lin
09	YSU	New S.A.S (14)	Lin
10	YSU	GD	Kessler (1)
11	YSU	GD	WRF 6-class (6)
12	YSU	GD	New Thompson (8)
13	YSU	GD	Stony Brook (13)
14	YSU	GD	WDM6 (16)

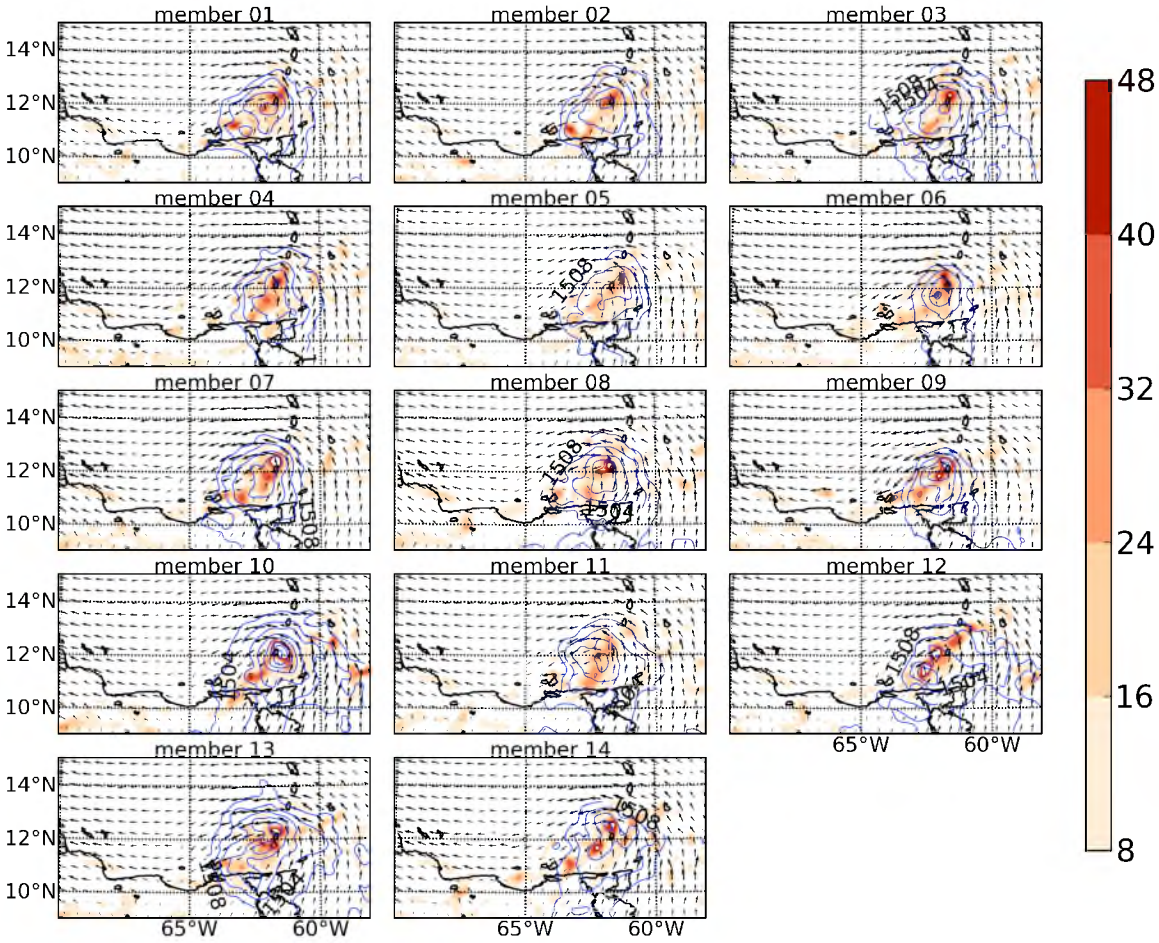

Figure 53. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of PHYS ensemble members at 850 hPa at 1800 UTC 24th.

Compared with the control ensemble (Figure 9) and many of the breeding ensembles (Figures 30-32), the PHYS ensemble shows relatively little intensity difference among the members at this time, especially in terms of vorticity. Compared with the ensembles where u and v were bred (Figure 30), the PHYS ensemble shows relatively little TC positioning diversity among the members. The TCs in the PHYS ensemble are well-organized at this time, however, and the genesis statistics are provided for this and comparable ensembles in Table 10. As expected from the synoptic overview, we see that many of the PHYS ensemble's members (12/14) show genesis at Ernesto's actual genesis time. While the spread among its members is low in terms of TC intensity and positioning, the PHYS ensemble produces genesis quite easily, which is likely because it is based around FNL ICs and BCs, whose deterministic forecast produced a strong TC itself. Recall that for comparison, Figure 51 provides an overview of this Ernesto deterministic forecast at 1800 UTC 24th. What we see in this figure is the fact that the storm is well formed by this time and that there is a vorticity maxima to the northeast of the geopotential height closed contours. Comparing this with the PHYS ensemble, we see that many of its members also produce similar intensity and positioning features at this time (Figure 53). Thus, we see that changing the physics parameterization schemes (Table 9) creates little diversity among the members in terms of spinning up Ernesto after 18 hours. Next, we will examine how this ensemble compares to the others in terms of error growth, error dimensionality, and error spatial spectra.

Error growth, dimensionality, and spread

Referring back to the ensemble bias (Figures 4-7), we see that in terms of vapor bias, the PHYS ensemble is always comparable to the other ensembles, but does particularly well at 700 hPa. In terms of θ bias, however, the PHYS ensemble

Table 10

Genesis statistics for

Hurricane Ernesto (forecast began 0000 UTC 24th)

Ensemble experiments	18Z 24th	12Z 25th
CNTL	07/14	13/14
PHYS	12/14	14/14
FNLq3	02/14	14/14
FNLuvqt3	01/14	12/14

consistently shows relatively low bias at 700 and 925 hPa, the latter of which shows a relatively great performance for model-error based ensembles in general. In addition, the PHYS ensemble's lower error at these levels occurs over much of the simulation period. While relatively low bias is a notable benefit of this multiphysics ensemble, it is important to compare this against the concurrent amount of spread produced, as helpful ensembles are expected to produce both accurate forecasts and also reasonable estimates of uncertainty.

Plotted in Figure 54 is the 850 hPa water vapor E-dimension and spread values for the CNTL, PHYS, and FNLq3 ensembles over time. These values are within 200 km radius of the corresponding TC center. Note the relatively slow E-dimension growth during the first 9 h in the PHYS ensemble compared to the CNTL and FNLq3. Relatedly, the midsimulation low dimensionality that arises for all ensembles is notably late in the PHYS ensemble, but arrives quite quickly with a sharper drop in dimensionality, compared to the other ensembles, after 18 h. This delayed drop in error dimensionality among the members of PHYS likely relates to the extra time it takes for the members to differentiate themselves due to their being based on model and not initial condition uncertainty. Despite these dimensionality differences, however, PHYS's vapor spread is within 0.1 g/kg of the FNLq3 ensemble over the entire period.

To determine differences in correlation between vapor dimensionality and spread for this set of ensembles, a 1000-sample bootstrap procedure is performed and the results are plotted in Figure 55. While the control ensemble does have some wide-ranging correlation samples below and near zero, at 500 cumulative samples, the PHYS ensembles show an overall weaker relationship between near-TC vapor dimensionality and spread compared with the CNTL and FNLq3 ensemble. FNLq3, by a notable margin,

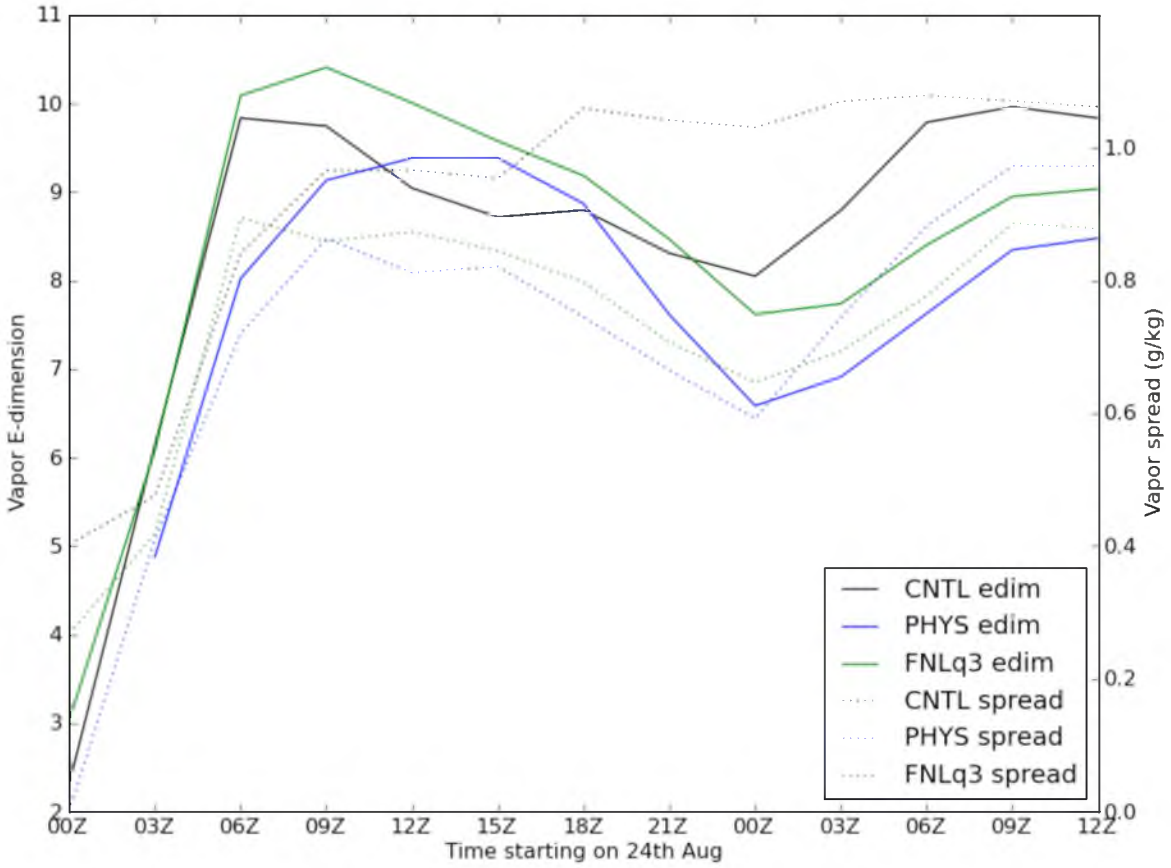

Figure 54. Time series of error E-dimension and spread (g/kg) for 850 hPa water vapor within 200 km radius of simulated TC center in the CNTL, PHYS, and FNLq3 ensembles.

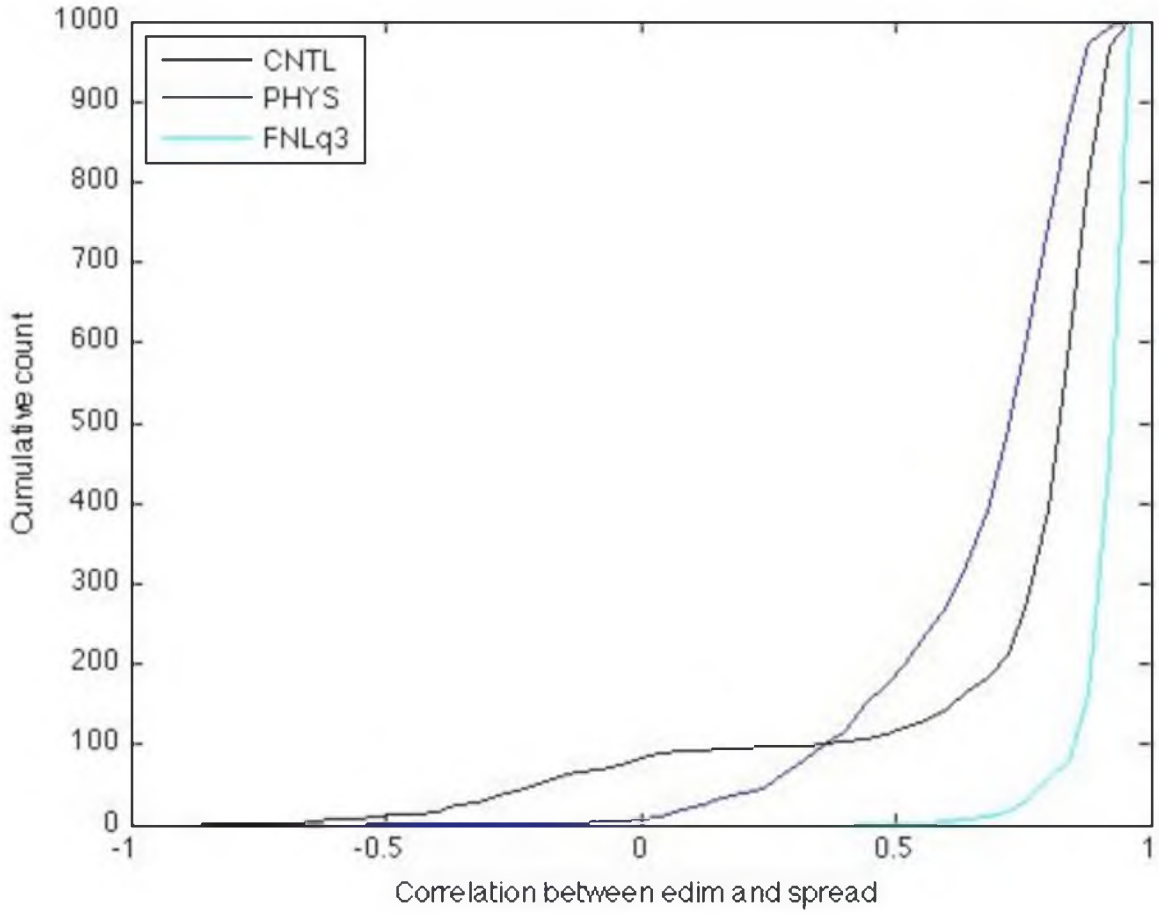

Figure 55. Cumulative distribution of bootstrapped correlation coefficient between E-dimension and spread based on 850 hPa water vapor within 200 km radius of simulated TC center over the entire simulation period of CNTL, PHYS, and FNLq3 ensembles. The two sets of values for each ensemble were bootstrapped for 1000 samples.

produces the highest correlation values of all three ensembles as, in general, it seems that breeding causes a notably strong relationship between an ensemble's error dimensionality and spread.

In order to view how the water vapor dimensionality of each of these ensembles changes as one approaches the TC center, the E-dimension metric is calculated for the time period of interest and plotted in Figure 56. Note that the typical pattern, of high dimensionality growth followed by decline around 1200 UTC 24th, occurs for each of these three ensembles in terms of vapor and θ and for almost all distances from the TC center. We first note that the relationship between near-TC and large radii dimensionality for the PHYS ensemble is more like that of the FNLth3 (Figure 24) and FNLq3 (Figure 56) ensembles compared with the FNLuvqt ensembles (Figure 36); we refer specifically here to θ and vapor patterns by radii, as wind displayed a similar profile over all ensembles. While the interpretation of this is difficult, it may mean that PHYS error growth, in terms of θ and vapor, at small versus large radii is closer to that of FNLq3 and FNLth3 than the ensembles where several variables are bred.

Note that the PHYS ensemble exhibits the least amount of dimensionality growth, compared to the CNTL and FNLq3 ensembles, during the first 6 h of the simulation. Note that this early reduced error dimensionality on the part of PHYS occurs fairly uniformly at all radii and for the three variables of interest. At the same time, we see that the heightened number of error variance directions resulting from the vapor breeding (Figure 56c,f,i) not only manifests itself in the vapor fields, but, after model spinup, in other variables as well. It is likely that these PHYS/FNLq3 differences result from the construction style of the breeding ensembles, which provide fully formed perturbations by 0000 UTC 24th, whereas for several hours after this, the PHYS ensemble members are only very slightly distinguished from each other. For Ernesto, this figure thus quantifies

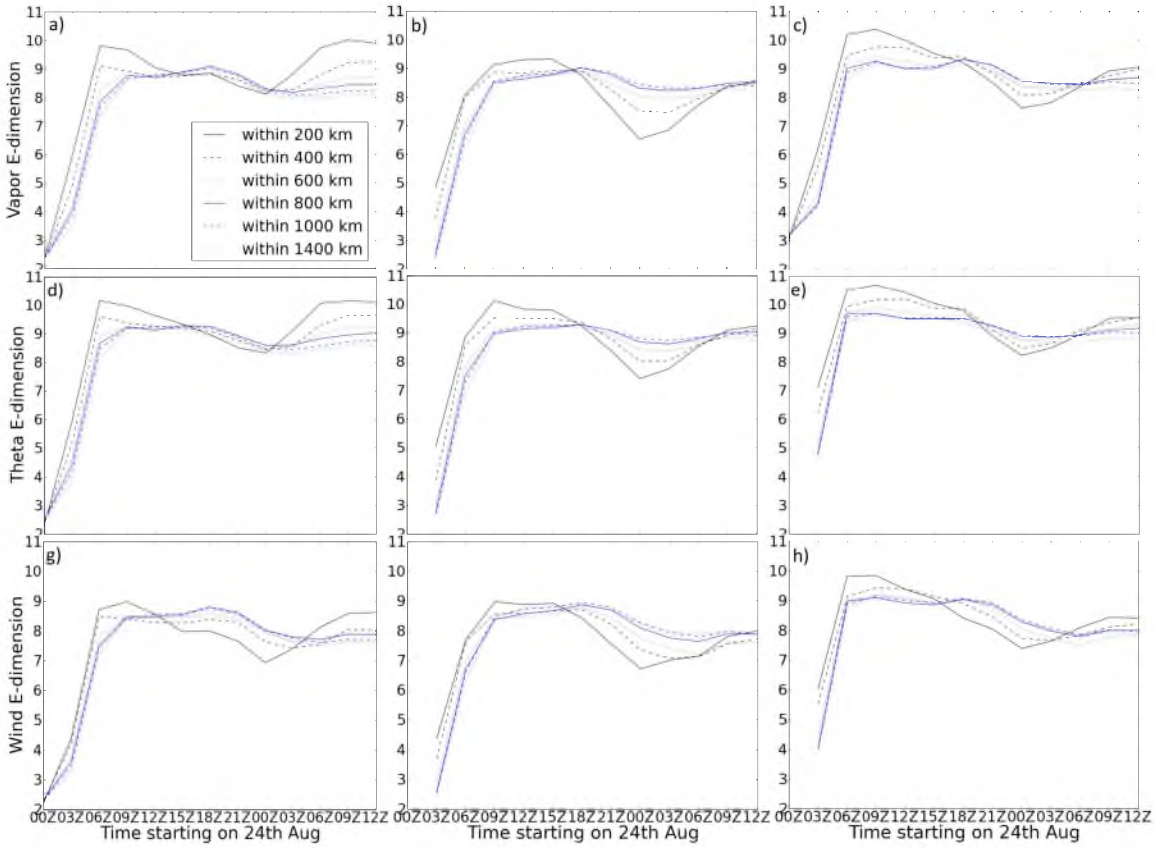

Figure 56. Time series of error E-dimension at 850 hPa and by radii from the simulated TC center for (a-c) water vapor, (d-f) theta, and (g-i) wind. The first column shows results for the CNTL ensemble, second column shows the PHYS ensemble, and the third column shows the FNLq3 ensemble.

the differences in error dimensionality between a simple IC-based perturbation method and a simple model error-based perturbation method. Finally, notice how, compared to the other variables, θ (Figure 56d-f) provides higher dimensionality over the first 9-12 h in all three ensembles. That this is likely related to the relatively heightened θ growth error values during the breeding period, compared to the other variables, is seen in the fact that the wind field has the slowest breeding error growth (Table 4; especially in terms of U) and also, quite consistently, the lowest early E-dimension values in all ensembles. Recall that while Kuhl et al. (2007) discussed how error growth leads to *lower* E-dimension in general, because of the way bred errors are incorporated into the final perturbations, *higher* error growth during the breeding period appears to lead to *higher* E-dimension values during the model spin-up and early simulation from 0000 to 1200 UTC 24th.

Again, essentially regardless of the ensemble or whether one examines vapor, θ, or wind, the number of notable error directions is significantly more variable near TCs than away from them. It consistently appears that faster dimensionality growth, near the TCs, over the first few hours of the simulation is accompanied by a faster decline in dimensionality after 1200 UTC 24th. In addition, these same values recover more quickly toward the end of the simulation from the temporary low-dimensional state. This was not the case generally in the work of Oczkowski et al. (2005), where fast declines into low-dimensionality did not usually mean a fast recovery to high-dimensionality (see their Figure 2). Our results do agree, however, in the fact that barotropic areas experience less transience, compared to baroclinic areas, in terms of ensemble dimensionality over time.

Early perturbation and error spatial spectra

In order to evaluate the spectral error differences between these ensembles, the FFT is used on the error of the 500 hPa geopotential height in the CNTL, PHYS, FNLq3, and FNLuvqt3 ensembles and the results are plotted in Figure 57. What we see is the fact that error spatial patterns are quite notable between the ensembles at 0300 UTC 24th, where the multiphysics ensemble has the highest error energy across most of the spectrum and the CNTL ensemble has the lowest. The bred ensemble's values are found roughly between those of the PHYS and CNTL ensembles at 0300 UTC 24th and these two converge with the PHYS ensemble at most parts of the spectrum by 3 h later (Figure 57b). While the control ensemble continues to show slightly lower error compared to the other ensembles at the large end of the spectrum at 1200 UTC (not shown), the more notable point is in terms of the PHYS ensemble. For example, while at 0300 and 0600 UTC it has notably higher error than the other ensembles, these differences largely disappear by 0900 and 1200 UTC. Thus, we see that when using an ensemble based around multiple parameterization schemes, one must be careful to note the prolonged ensemble spin-up period as compared to ensembles based around initial uncertainty. It is interesting that, despite the fact that the breeding schemes were originally (i.e., before the breeding period) based on the same type of ICs as the control ensemble, these breeding ensembles apparently share more in common with the PHYS ensemble in terms of the early establishment of error structure across the spectrum (Figure 57b).

To determine how well the PHYS and other ensembles are creating perturbations that simulate error at different parts of the spectrum, we again use an FFT calculation and plot in Figure 58 the perturbation spectrum for 500 hPa geopotential height in the CNTL, PHYS, FNLq3, and FNLuvqt3 ensembles near the start time. Note how, unlike its error

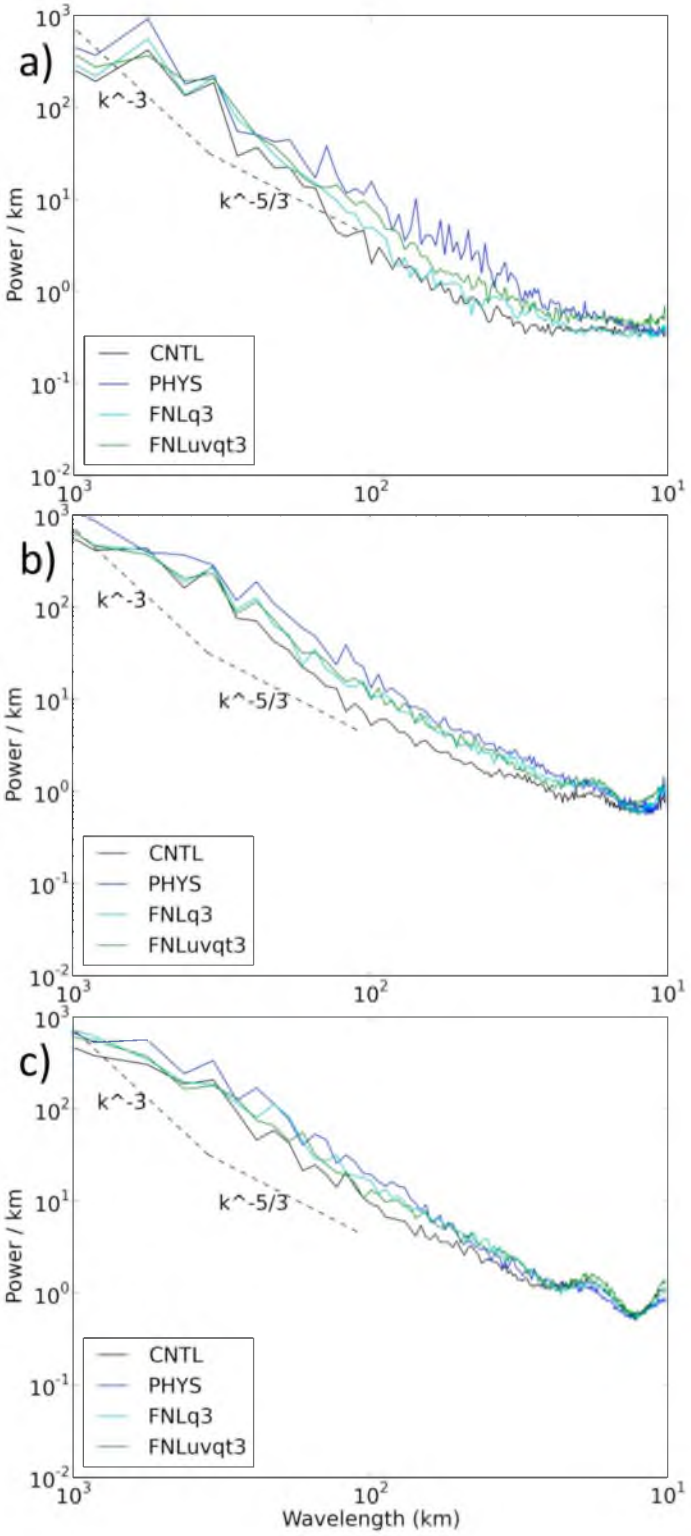

Figure 57. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height error across a zonal and ensemble average for CNTL, PHYS, FNLq3, and FNLuvqt3 ensembles at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

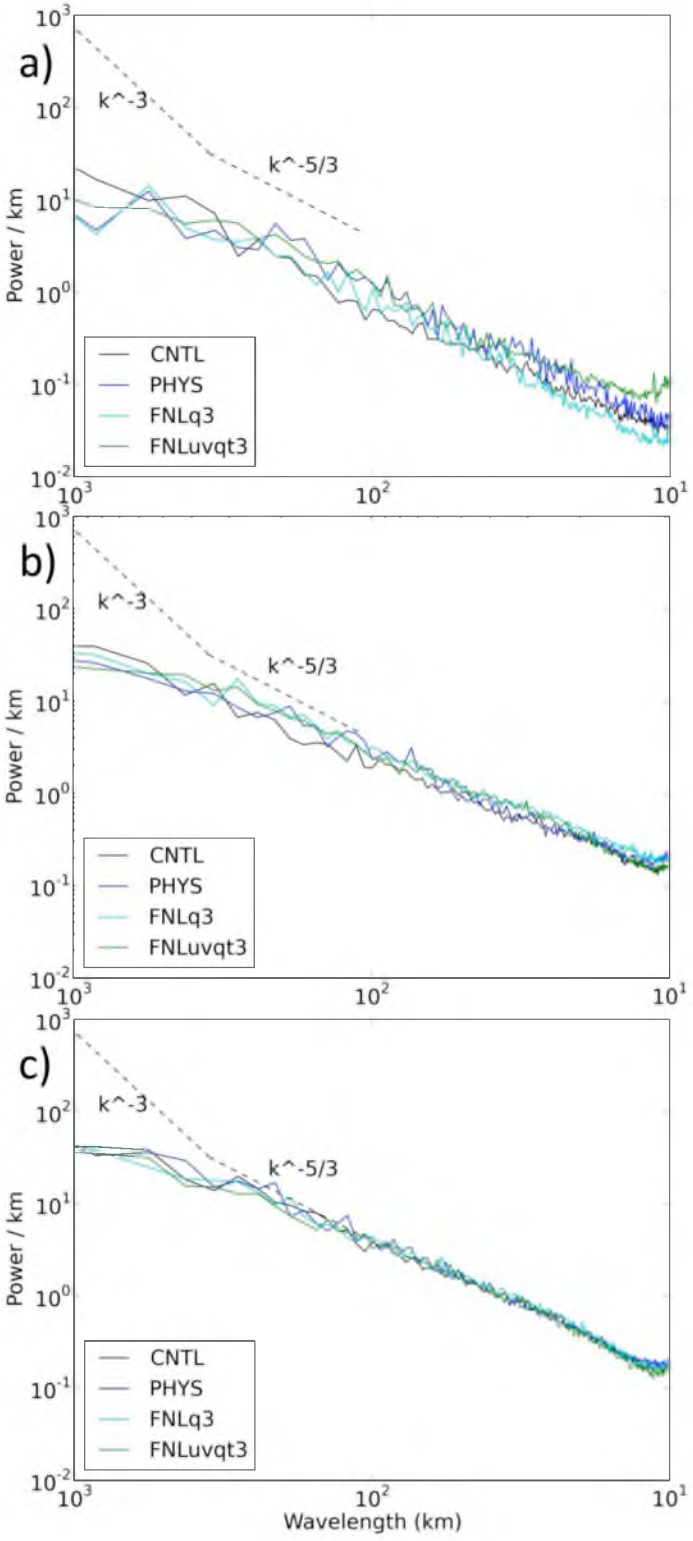

Figure 58. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height perturbation from the ensemble mean, then across a zonal and ensemble average, for CNTL, PHYS, FNLq3, and FNLuvqt3 ensembles at a) 0300 UTC, b) 0600 UTC, and c) 0900 UTC 24th August.

pattern, in Figure 58a, the PHYS ensemble does not show notably high perturbation energy, compared with the other ensembles, across the spectrum. Thus, considering perturbation and error energy, the PHYS ensemble appears to be slightly underdispersive compared to the other three ensembles, which makes sense considering its mode of implementation. Also noteworthy is the fact that the control ensemble shows particularly high dispersion compared to the breeding and PHYS ensembles due to its relatively low error at most parts of the spectrum. Overall, we see how remarkably similar the disparate ensembles are in terms of error and perturbation spectra, except for slight under-dispersion by the PHYS ensemble at the start of the simulation and the low error of the CNTL ensemble around 100 km present through 0900 UTC (Figure 57c).

Discussion

Overall, the ensemble based on various physics parameterizations produces relatively little variance between TCs in terms of intensity or spatial positioning. Most of the PHYS members show a well-formed storm by best track genesis time and, in fact, have the easiest time spinning up TCs compared to any of our other ensembles. We believe this is largely explained by the fact that the ICs and BCs behind the PHYS ensemble, in our test deterministic simulation, produce a storm very similar to those of the PHYS ensemble members. Essentially, the multiphysics ensemble demonstrates little uncertainty in the forecast, especially compared to the CNTL or FNLuvqt3 ensembles. While only many more cases would demonstrate this guideline to statistical significance, in Ernesto, the PHYS ensemble appears to have notably less TC intensity dispersion compared to ensembles based on more sophisticated constructions.

In general, the PHYS ensemble produces similar bias compared to the other ensembles, in terms of vapor and θ over time. It also shows lower TC-proximate (Figure

54) 850 hPa vapor dimensionality and spread throughout the simulation period compared with both the CNTL and FNLq3 ensembles. Relatedly, upon examining the PHYS ensemble's bootstrapped correlation between vapor spread and dimensionality, we find that it is also notably lower than that of the CNTL and especially lower than that of the FNLq3 ensembles. While the process is not entirely clear, both of these facts likely relate to the particular ensemble creation details of the breeding and Ensemble Transform (ET) scheme and the particular way in which they translate sampled error growth rates and directions into perturbation patterns across their members.

In particular, compared with the other ensembles of interest, the PHYS ensemble shows slower dimensionality growth in terms of θ and vapor, at the start of the simulation. Subsequently, PHYS shows a relatively late midsimulation decline in dimensionality. It looks as if the slow PHYS dimensionality changes at the beginning of the simulation are due to the fact that the ensemble differences are based on the slow-working parameterization differences, rather than initial condition differences present in CNTL and FNLq3. Similar to the other ensembles, the PHYS ensemble shows notably higher dimensionality close to TCs compared with those values at large radii, which reflects enhanced local baroclinic energy conversion processes taking place (Oczkowski et al. 2005).

In terms of 500 hPa geopotential spatial spectral pattern, the PHYS ensemble takes about 6 h of simulation time for its error energy to decrease to almost that of the CNTL and FNLq3 ensembles. After this time, PHYS shows slightly more error than the breeding and CNTL ensembles along with similar perturbation energy. Thus, the PHYS ensemble shows slightly more underdispersion in terms of 500 hPa geopotential height compared to the breeding ensembles, and is notably more underdispersive compared with the CNTL ensemble.

Highlights of Chapter 6

- PHYS ensemble produces relatively little variance among the members but strong TCs, which is largely because its members do not vary significantly from deterministic simulation.
- PHYS shows notably higher error compared to CNTL and breeding ensembles, over much of the spectrum, for the first 6 to 9 hours of simulation time.
- Compared to the other ensembles, PHYS is markedly underdispersive over this time period.
- Similar to the ensembles based on initial perturbations, the PHYS ensemble shows notably higher dimensionality at high versus low radii from the TC.
- Likely because of this difference in construction between PHYS and FNLq3 (which has initial perturbations), there a much stronger relationship between dimensionality and spread in the latter versus the former.

CHAPTER 7

IMPLICATIONS OF THE ENSEMBLE FORECASTS: IMPACT OF ENVIRONMENTAL PRECURSORS ON TC GENESIS

In order to better understand the relative impact of our various ensemble results on TC genesis, we move now from more abstract statistics to ones that relate to TC genesis processes themselves. Essentially, instead of considering ensemble error dimensionality and spatial spectra, we will now move toward examining the impact of low level water vapor, environmental wind shear, and other factors that likely play a role in determining the genesis statistics for each particular ensemble. This way we can more effectively link model error growth and characteristics with eventual genesis outcomes.

We will first do this by using our full complement of Ernesto forecast ensembles instead of examining the peculiarities of certain ensembles in isolation. In total, we have 11 ensembles with relatively unique initial conditions. The PHYS ensemble, due to its physics-based perturbations and nonunique initial conditions, is excluded. So, in all, we have 154 members forecasting the proto-Ernesto. While each member is unique at 0000 UTC 24th August in some regard, we will be careful to provide an accurate estimate of statistical significance in light of the fact that our initial number of degrees of freedom, in terms of a particular field, will certainly be fewer than the 154 members in the sample. Degrees of freedom here means the number of values in the sample that are free to vary. The members are separated into genesis and nongenesis bins, depending on whether they

show genesis at the best track genesis time of 1800 UTC 24th August. Again, genesis is defined here as occurring if the forecast contains a MSLP contour of 1009 hPa or below and a closed 850 hPa geopotential contour of 1496 m or below, both of which must be found within, or overlap, a closed wind circulation at 850 hPa.

The relative operating characteristic curves

To begin this analysis, we will use the relative operating characteristic (ROC) curve to examine the effect of certain values of vertical shear, water vapor, and other metrics on genesis likelihood. This statistic and plot is here created by plotting the fraction of true positives out of the positive genesis cases, versus the fraction of false positives out of the number of negative (or null) genesis members. The members are binned in each category depending on their initial values of a particular field relative to the threshold of choice for that field, and whether the same member subsequently predicts genesis. In statistics generally, this true positive rate is also known as the sensitivity and the false positive rate is one minus the specificity. A contingency table is created by using incremental thresholds of the chosen environmental data and whether genesis, in the members corresponding to those particular thresholds, occurs or not. By way of plot orientation, a perfect predictive factor would result in a point in the upper-left hand corner, representing no false negatives and no false positives. By contrast, perfectly random data, or data in which there is no predictive power, will lead to a "no-discrimination" line that runs diagonal from the bottom left to the top right corner. Positive area under the ROC curve, and above the no-discrimination line, is indicative of positive predictive power. While positive area would seem a misnomer, it is used here to differentiate positive area above the no-discrimination line with negative area below the random line.

The impact of initial water vapor on genesis likelihood

Because of its impact on convection, one of the most common environmental precursors of TC genesis is mid-to-lower-level water vapor (Bister and Emanuel 1997; Sippel and Zhang 2008). The concentration of the near-TC moisture has long-been recognized as being an important predecessor of genesis (Gray 1968), and, because high values of water vapor are found extensively throughout the Tropics, is seen as allowing, rather than necessarily causing genesis to occur. Seeing as how water vapor appeared to have a notable impact on genesis across our various ensembles, this is the first field we study in terms of its genesis predictive power. Plotted in Figure 59 are four ROC plots, which display the strength of the relationship between TC genesis at 1800 UTC 24th August and water vapor within 100 km of the TC center at 500, 600, 700, and 850 hPa at 0000 UTC 24th. Also plotted is the "no-discrimination" diagonal line, where no predictive power would be present. First notice the fact that the vapor pattern at the start time and at 500 hPa shows a small amount of positive area under the ROC curve, indicating that water vapor in this region, at the start of the simulation, has a higher true positive rate than false positive rate. This is to say that there is some positive predictive power present. As for the construction of the curve, threshold values for Figure 59a are incremented by 0.5 g/kg between 0 and 7 g/kg.

Notice that at 600 hPa (Figure 59b), there is more positive area under the curve than at 500 hPa. The 700 hPa ROC figure shows more positive area still, and finally at 850 hPa (Figure 59c,d), positive predictive power in terms of water vapor is maximized and the effect then declines in power below this level. Focusing in on the ROC curves at 700 and 850 hPa, it is quite impressive how well water vapor within 100 km of the nascent

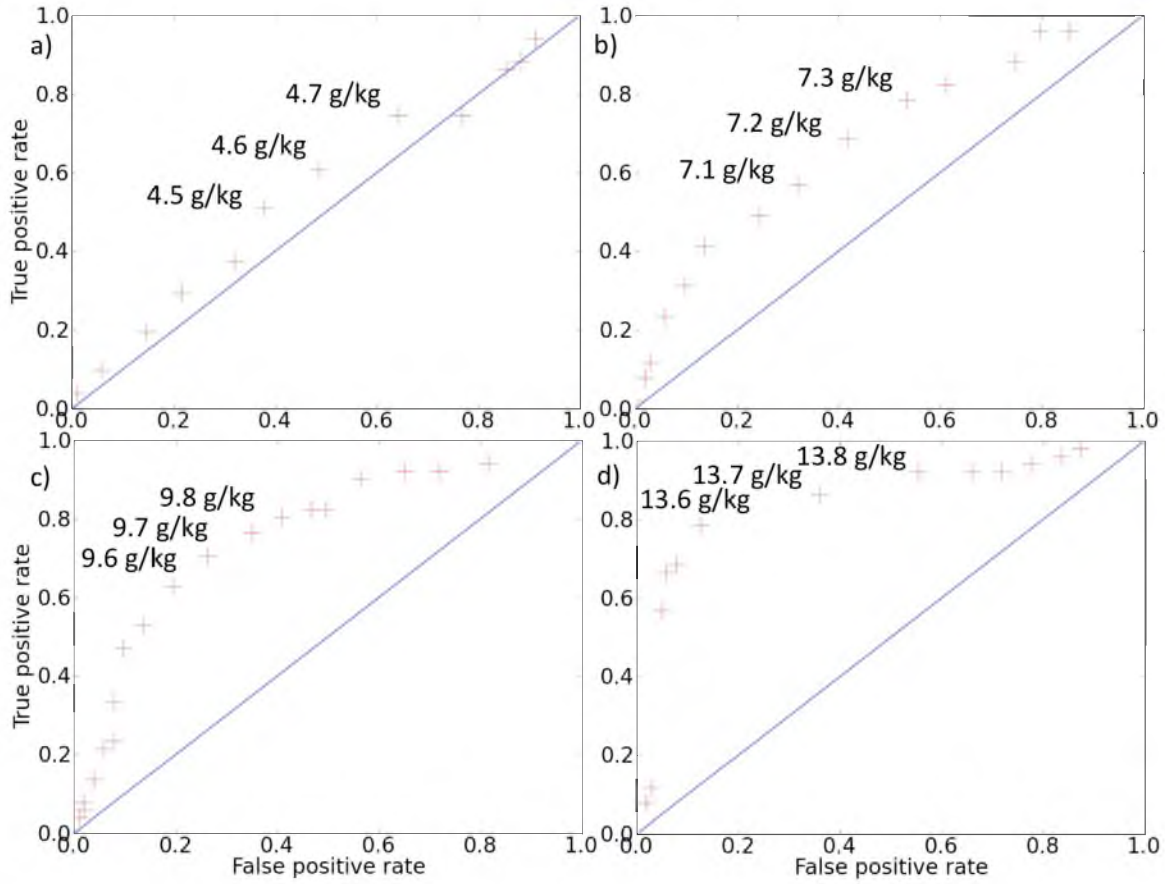

Figure 59. Receiver operating characteristic (ROC) curves showing relationship between genesis at 1800 UTC 24th August and water vapor within 100 km radius of simulated TC center a) at 500 hPa, b) 600 hPa, c) 700 hPa, and d) 850 hPa at 0000 UTC 24th August.

TC center predicts TC genesis 18 h later. This is especially notable because of the fact that vapor is often seen as generally being sufficient across most of the Tropics during the summer months and that a pre-existing vorticity maximum and so-called stochastic processes (Simpson et al. 1997) are really the drivers behind genesis. Examining the data behind Figure 59d, we see that a near-TC water vapor threshold of 13.8 g/kg at 850 hPa produced 40 true positive, 13 false positive, 11 false negative, and 90 true negative genesis cases 18 h later. Thus, with this 13.8 g/kg initial criteria, we were able to accurately predict 78% of the genesis cases and 87% of the null cases. By all accounts, these type of numbers are indicative of an important physical relationship. Compared to its usual role in the Tropics as a necessary but not sufficient condition for genesis, water vapor in our ensembles appears to play a more active part in genesis occurrence.

To illustrate how the water vapor composites differ for the genesis and nongenesis cases, these are constructed and presented in Figure 60 for the 500 and 600 hPa levels, and in Figure 61 for the 700 and 850 hPa levels. First, notice the degree to which water vapor levels shift from being associated with the easterly wave at 700 hPa, compared with its much more localized distribution at 850 hPa. Focusing on the differences between the genesis and nongenesis composites of water vapor at 500 hPa (Figure 60a), notice the fact that the genesis composite has roughly 0.3 g/kg more water vapor near the TC center and that this difference extends for a few hundred square kilometers in any direction. Recall that at 500 hPa, a 0.3 g/kg vapor difference is not trivial, especially when it occurs so near the pre-Ernesto disturbance. This difference is quite localized, however, as the 500 hPa genesis composite shows a notably smaller amount of water vapor in the center of the domain. At 600 hPa (Figure 60b), we see that the difference, between the composites, near the TC (notice the X) is now up to near 0.6 g/kg. Again,

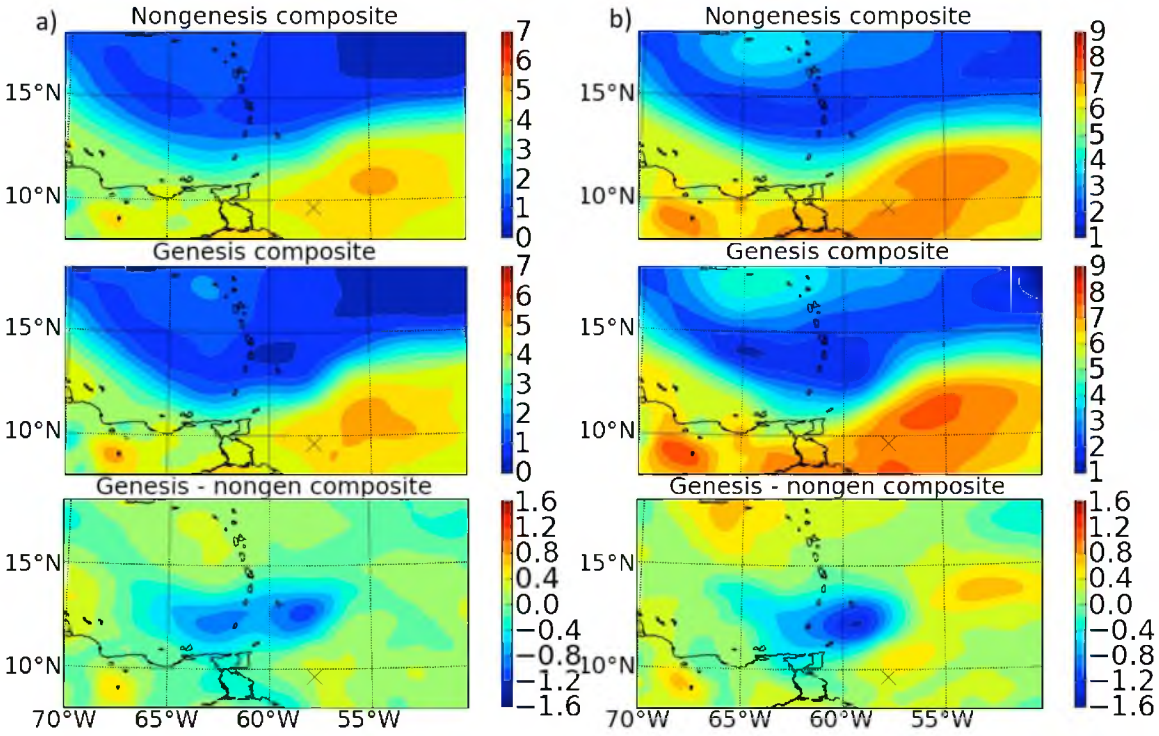

Figure 60. Genesis and nongenesis composites at 0000 UTC 24th August in terms of water vapor (g/kg) at a) 500 hPa and b) 600 hPa. X marks the location of the pre-Ernesto disturbance.

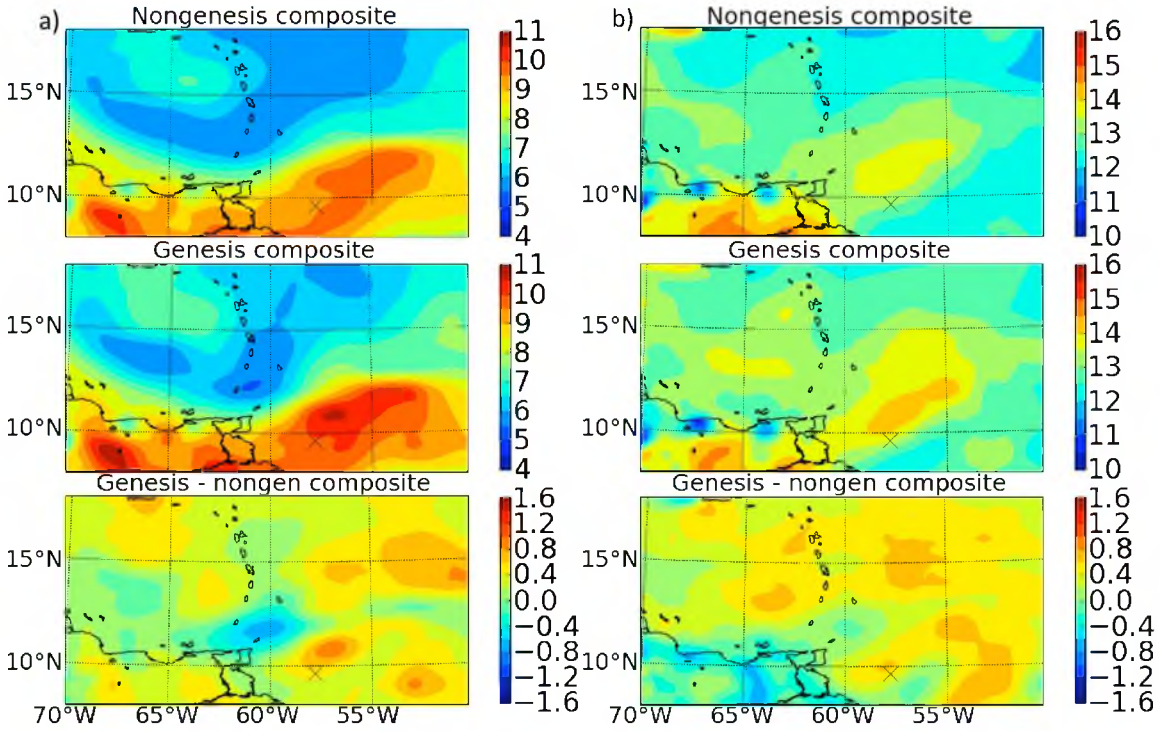

Figure 61. Genesis and nongenesis composites at 0000 UTC 24th August in terms of water vapor (g/kg) at a) 700 hPa and b) 850 hPa. X marks the location of the pre-Ernesto disturbance.

notice that in the center of the domain the genesis composite shows less water vapor than the nongenesis composite, due to the fact that the genesis cases tend to accompany a higher-amplitude wave, which produces a more southward extension of the midlevel dry tongue seen across the northern half of the domain. Moving to 700 and 850 hPa (Figure 61), we see that there is an anomalously high amount of water vapor for at least a hundred kilometers radius from the TC in the genesis cases. In fact, at the genesis-important level of 850 hPa, the difference between the composites reaches 0.8 g/kg nearly surrounding the nascent TC center.

The impact of initial vertical shear on genesis likelihood

Next to moisture, one of the most important environmental precursors for genesis is the prevalence of low vertical wind shear. Again, as with water vapor, and because of the frequency with which it occurs, low shear is often seen as a requirement, rather than an driver of genesis. Nevertheless, and to determine the actual nature of this interaction in Ernesto, using our 154 Ernesto forecasts, we next examine the effect of early-simulation vertical wind shear on TC formation at 1800 UTC 24th August. We note here that some ensembles may experience only slightly unique wind profiles, and only after 0000 UTC, because not all ensembles specifically created unique initial wind perturbations. This, however, will be more fully addressed in the statistical significance chapter to follow. Plotted in Figure 62 are the ROC curves for the impact of vertical wind shear values within 500 km of the TC center at 0000, 0300, and 0600 UTC 24th August on genesis 18 h later. Vertical shear is here calculated by subtracting the components of the wind at 200 hPa from those at 850 hPa (Frank and Ritchie 2001; Knaff et al. 2004). What we find is indeed a strong, negative impact of vertical shear at the start of the simulation period

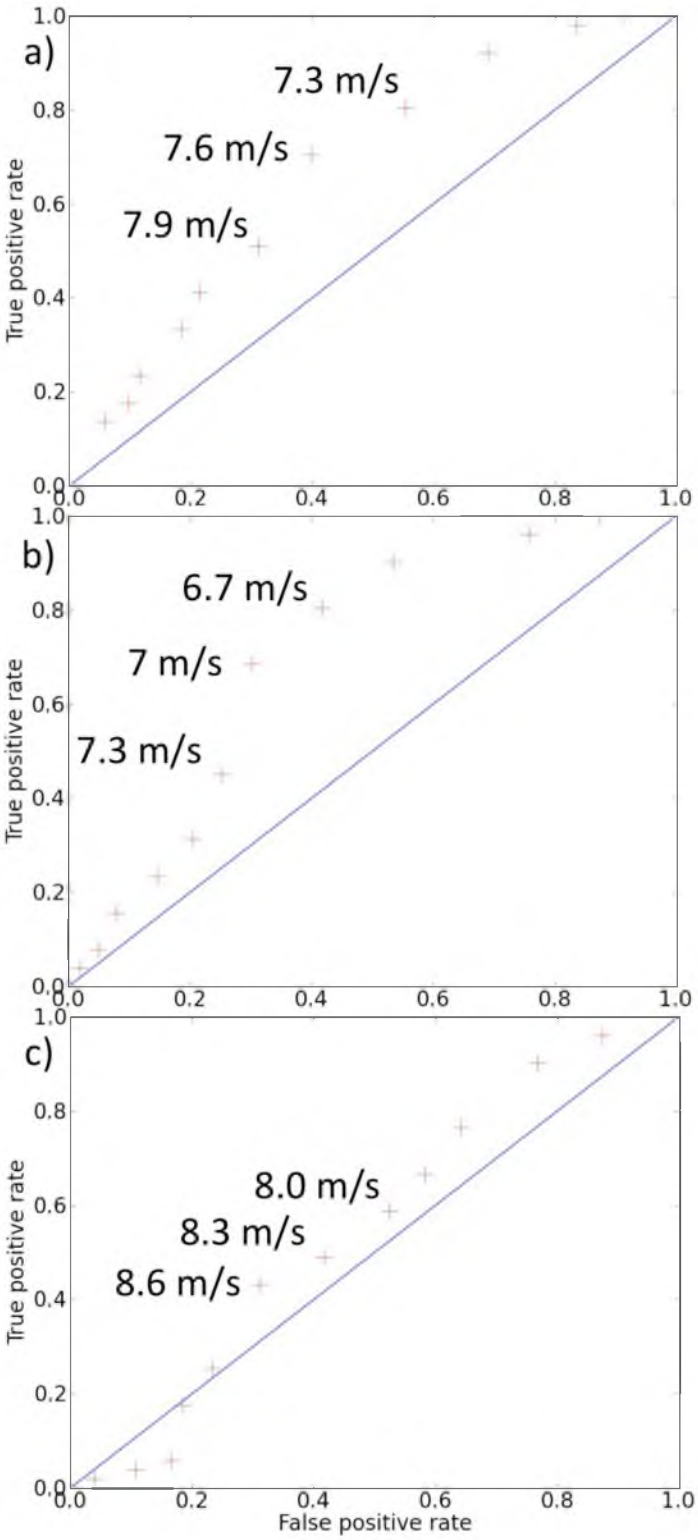

Figure 62. ROC curves showing relationship between genesis at 1800 UTC 24th August and vertical shear within 500 km radius of simulated TC center a) at 0000 UTC, b) 0300 UTC, and 0600 UTC 24th August.

on genesis 18 h later (Figure 62a). Despite this notable, initial impact, the effect of shear at 0300 UTC is even greater (Figure 62b). Note the large negative area portraying the vertical shear-genesis relationship. In terms of what this ROC curve is specifically describing, for a threshold of 7.6 m/s shear averaged within the 500 km radius, we find that this leads to 35 true positive, 31 false positive, 16 false negative, and 72 true negative genesis predictions. By way of interpretation, and considering our 7.6 m/s criteria, of the 103 null cases, 72 of these (or 70%) were predicted; of the 51 genesis cases which occurred, 35 (or 69%) were correctly predicted.

Similar calculations place the curve well below the no-value diagonal line at 0600 UTC. Note, however, that at this time, we see that vertical shear is not as powerful a predictor, as the area within the curve is notably smaller. This likely relates to the fact that TCs themselves create shear because of the way an anticyclone becomes established above the lower-level cyclone; since we did not remove the vortex in these ROC calculations and instead averaged within a 500 km radius, the mixed signal at 0600 UTC (Figure 62c) may relate to the fact that, starting at this point, vertical shear accompanies the presence of TCs instead of inhibiting their presence. Indeed, the slightly positive relationship between shear at 0900 UTC (not shown) and genesis confirms this trend.

The impact of pre-existing vorticity on genesis likelihood

Contrary to the water vapor and vertical shear amounts, which are seen as being necessary but not sufficient for genesis, pre-existing vorticity maxima are often viewed as being the impetus for genesis. There has been much discussion as to whether this pre-existing vorticity most enables genesis when it arises first in the mid- or lower levels, which two scenarios are often described as a top-down (Bister and Emanuel 1997; Ritchie and Holland 1997; Simpson et al 1997) or bottom-up (Davis and Bosart 2001;

Hendricks et al. 2004; Reasor et al. 2005; Montgomery et al. 2006) path to genesis, respectively. To evaluate the impact of pre-existing vorticity levels on forecasts of TC genesis, in Figure 63, we plot a ROC curve of the relationship between lower level (8-900 hPa) and midlevel (5-600 hPa) vorticity at 0000 UTC 24th on TC genesis 18 h later. What we find is that lower-level vorticity has a weak, but positive genesis predictive power. This power, or area under the curve, is not manifest at all thresholds, but mostly occurs at 100 km-radius-average values of 8, 8.5, and 9 x 10^{-5} s^{-1}. In fact, when using the predictive abilities of the 8.5 * 10^{-5} s^{-1} threshold, we find that of the 51 genesis cases which occurred, 33 (or 65%) were correctly predicted; for the 103 null cases, 55 (or 53%) were predicted. Considering the tradeoff that is made between true positive rate and false positive rate when considering thresholds, the fact that we are able to predict 65% of the genesis cases while not falling below 50% accurate prediction of the null cases demonstrates the small but positive signal in the relationship between early low-level vorticity and genesis.

While initial lower-level vorticity was found to positively impact genesis, that of the midlevels did not have such predictive power. In fact, as seen in Figure 63b, midlevel vorticity had a mixed if slightly negative relationship with genesis at 1800 UTC 24th Aug. While threshold vorticity values of 2 and 6.5-8 x 10^{-5} s^{-1} shows a positive relationship with genesis, the correlation is negative when applying thresholds of 2-6.5 x 10^{-5} s^{-1}. That lower levels of midlevel vorticity would actually inhibit TC development is something we did not anticipate. Such a negative relationship between initial midlevel vorticity and genesis has not been notably explained in the literature and in fact may be an artifact of only examining forecasts of Ernesto. While we will later look at the statistical significance of these relationships, we first break down the genesis and nongenesis cases

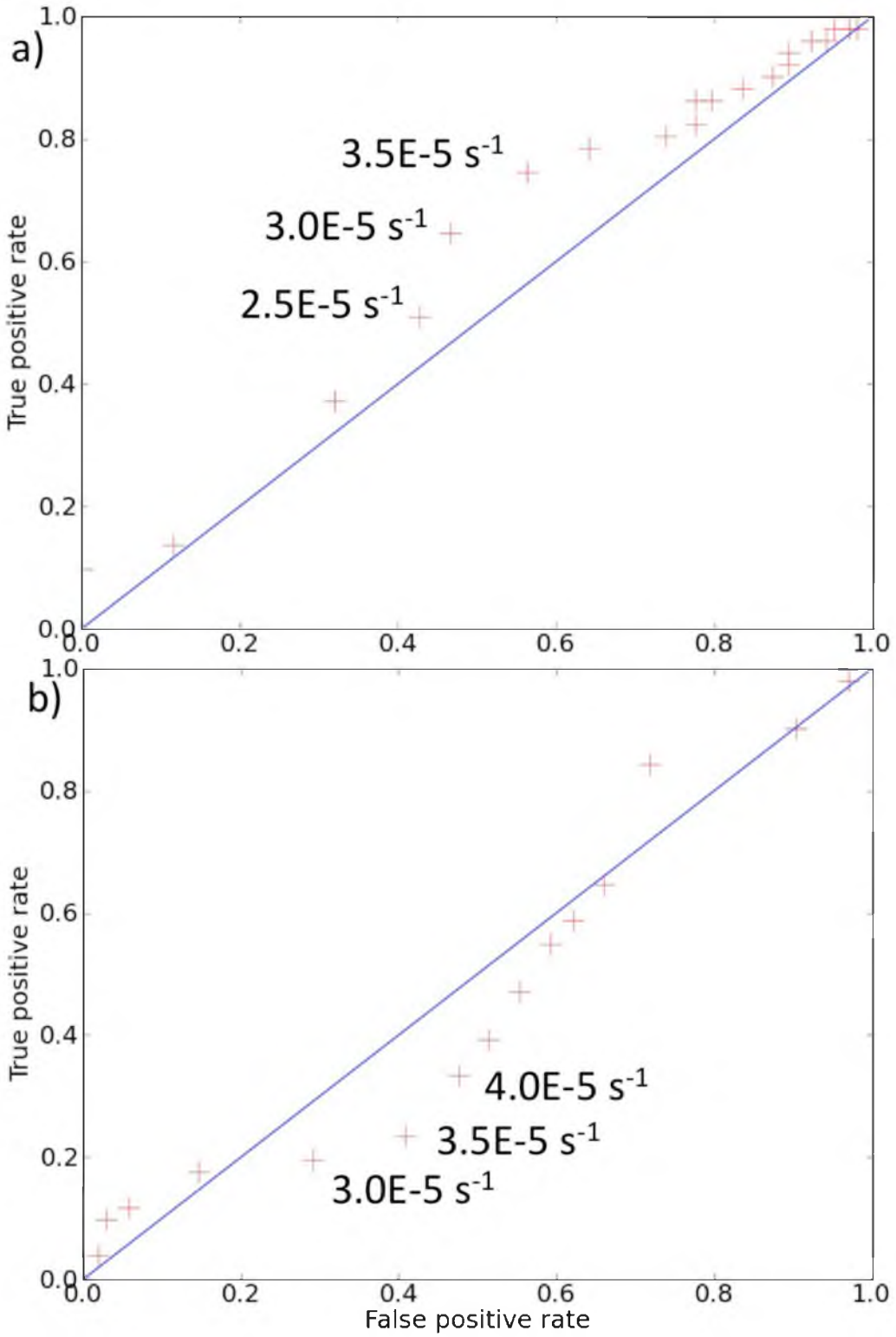

Figure 63. ROC curves showing relationship between genesis at 1800 UTC 24th August and vorticity within 100 km radius of simulated TC center a) from 8-900 hPa and b) from 5-600 hPa at 0000 UTC 24th August.

into composites of both lower and midlevel vorticity and plot these in Figure 64. What we see in part a) is that the genesis cases do indeed start with a more notable vorticity maxima associated with the lower level easterly wave. Note that because this lower-level wave is weaker in the nongenesis case, the magnitude of the vorticity over the inverted wave downstream is also weaker. Since vorticity is negative in this down-stream portion of the wave, the nongenesis composite has smaller negative values than the genesis composite, thus leading to the negative *difference* values directly south of the Lesser Antilles in Figure 64a. Examining the midlevel vorticity composites (Figure 64b), we notice that the maximum positive *difference* vorticity values are quite similar at this height, except the shift between the positive and negative difference values is now oriented in the north-south direction. These differences from 5-600 hPa, however, appear to be caused by the fact that the nongenesis case shows higher vorticity values from 55 to 60°W and between 10 to 13°N. Again, the fact that these vorticity differences at 5-600 hPa do not significantly affect genesis in a conclusive manner is seen in the ROC curve of Figure 63b.

In order to examine the impact of the strength of the 850 hPa easterly wave on genesis 18 h later, 850 hPa winds for genesis and nongenesis composites are calculated at 0000 and 0300 UTC 24th and displayed in Figure 65. What we see in these four subfigures is that the easterly wave is positioned around 55°W and 12°N. First, notice in part a) the way that the wave has a much more developed circulation in the genesis compared to the nongenesis composite. In the nongenesis composite at this time, a northerly component of the wind on the downwind side of the wave axis is conspicuously absent compared to the genesis composite. Specifically, in the genesis cases, winds around 58°W and 12°N are 2-4 m/s stronger, and with much more of a northerly (i.e., vortical) component compared with the nongenesis members. These

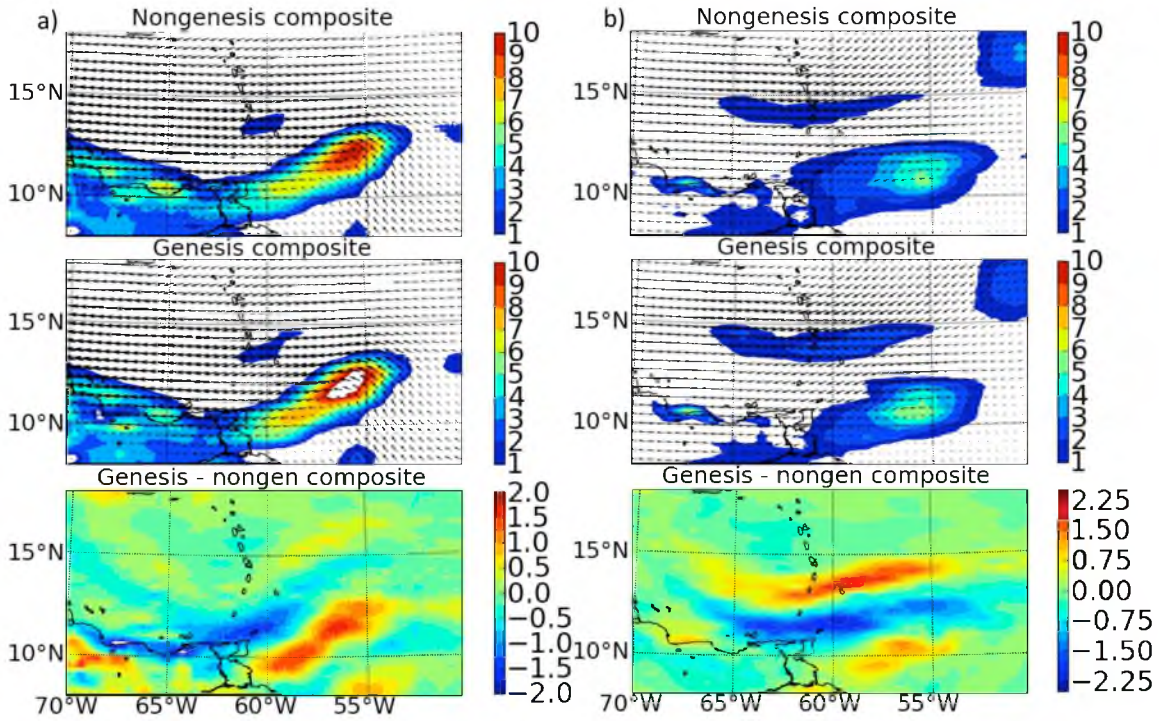

Figure 64. Genesis and nongenesis composites at 0000 UTC 24th August in terms of vorticity (s^{-1}) and the wind field at a) 8-900 hPa and b) 5-600 hPa.

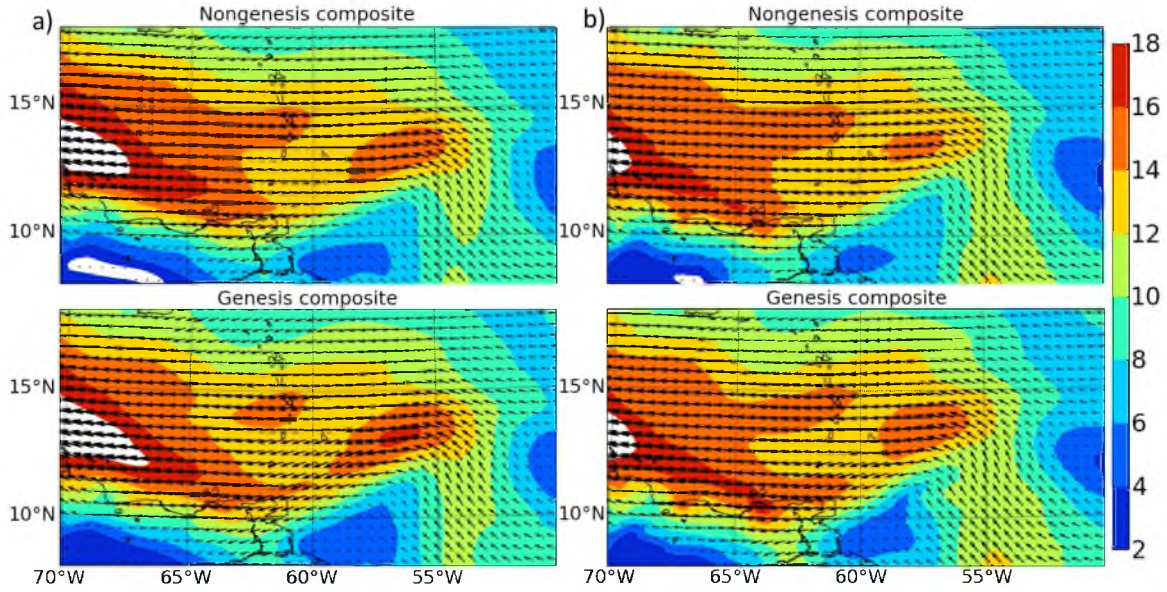

Figure 65. Genesis and nongenesis composites in terms of 850 hPa wind speed (m/s) and circulation at a) 0000 UTC and b) 0300 UTC 24th August.

directional differences are even more prominent at 0300 UTC (Figure 65b). Recall that the impact of these differences are reflected in the genesis result *15-18 h later*, despite the other complicating environmental factors present. While the genesis process is often seen as stochastic even in the presence of a pre-existing wave (Simpson et al. 1997; Reasor et al. 2005), in this study, we find that TC genesis likelihood is very much related to the strength of the pre-existing low-level wave up to 18 h before genesis (the statistical significance of these calculations will be discussed below).

The impact of upper-level warmth on genesis likelihood

While upper-level warmth is not generally seen as necessarily causing or inhibiting genesis, especially compared to SSTs, water vapor, shear, and vorticity, it nevertheless plays an integral role in the genesis process. Because of this and our large number of ensembles forecasting the genesis of Hurricane Ernesto (2006), we want to determine the relationship, if any, between initial levels of core potential temperature and genesis at 1800 UTC 24th. Using our 11 ensembles with unique initial conditions, ROC curves are calculated and plotted in Figure 66. We find that initial upper-level θ values do not have a significant impact on genesis statistics. This is seen in the erratic behavior of the 0000 UTC ROC curve, which switches between a positive and negative genesis likelihood, depending on the θ threshold. This lack of a strong signal makes sense, as at a particular height, large-scale θ values in the Tropics are generally similar over broad swaths of ocean. Move forward 3 h in our simulations, however, and the picture changes significantly. Notice how in Figure 66b, effect of upper-level θ on genesis is positive and consistent across various core θ thresholds. This is surprising, as we did not expect slight upper-level warmth to encourage genesis, nor did we expect notable warm

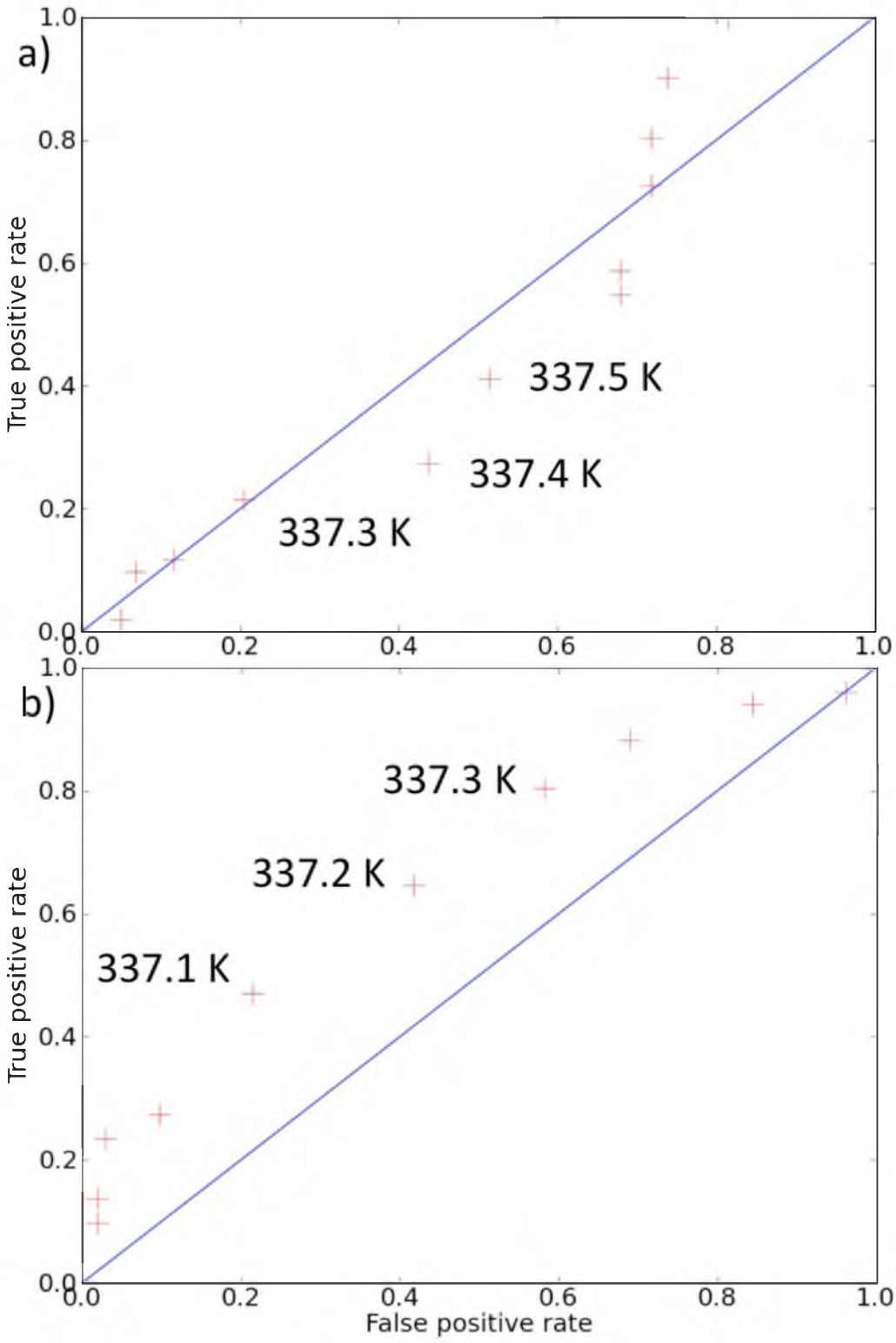

Figure 66. ROC curves showing relationship between genesis at 1800 UTC 24th August and theta within 100 km radius of simulated TC center from 2-500 hPa at a) 0000 and b) 0300 UTC 24th August.

cores to be present 15 h before best track genesis. In order to view the extent of these warm cores and how they varied between the genesis and nongenesis cases, the two corresponding composites are created and presented in Figure 67. What we see is that at the initial time (Figure 67a), there is only a slight increase (i.e., 0.1K) of upper-level θ from the nongenesis to the genesis composite. Three hours later, however, the genesis composite shows values around 1.0 K greater than the nongenesis composite in the pregenesis region, which is southeast of the Lesser Antilles. This demonstrates not only that the members showing genesis possess a notable warm core, but that many of these members possess genesis-like characteristics 15 h before genesis. In other words, whether individual member will forecast genesis or not, in our forecasts of Ernesto, is often decided well before most of the forecast has been made. Considering the often stochastic nature of genesis (Simpson et al. 1997; Reasor et al. 2005), this result was not expected.

Statistical significance of environmental precursor impacts on genesis

In order to investigate the statistical significance of the TC genesis relationships which have been enumerated, we use the Kolmogorov-Smirnov (KS) two-sample test. This nonparametric test provides a p-value, which describes the odds of the two samples being drawn from the same continuous distribution. Presented in Table 11 are the results of the KS test being used on various of the above-mentioned environmental precursors of genesis, using their genesis and nongenesis composites; see Table 2 in conjunction. In order to account for the fact that the 154 members may actually represent many fewer degrees of freedom in a particular field, we split the sample into groups of two ensembles each and determine the p-value for the particular genesis precursor relationship in each of

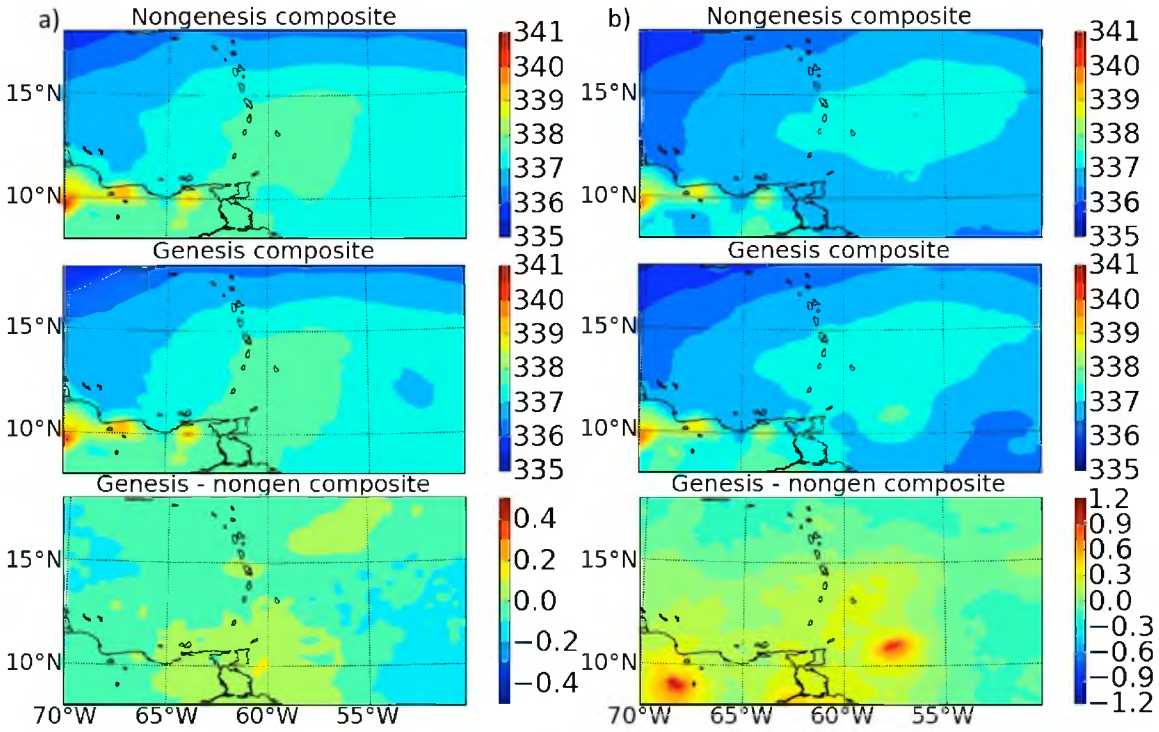

Figure 67. Genesis and nongenesis composites in terms of theta (K) from 2-500 hPa at a) 0000 UTC and b) 0300 UTC 24th August.

Table 11

P-values from the Kolmogorov-Smirnov test performed between

genesis and nongenesis composites of various parameters in ensemble groupings

Field	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8	8-9	9-10	Med	0-10
Shear00Z	.02	.27	.49	.32	.11	.02	.25	.85	.00	.02	.18	.00
Shear03Z	.01	.00	.04	.57	.08	.02	.35	.06	.00	.02	.03	.00
Shear06Z	.05	.04	.05	.10	.19	.11	.87	.13	.00	.29	.11	.11
Shear09Z	.19	.54	.19	.10	.25	.04	.99	.18	.00	.62	.19	.24
Vapor500	.94	.93	.04	.00	.00	.02	.03	.21	.49	.49	.13	.39
Vapor600	1.0	.17	.01	.00	.00	.00	.01	.03	.12	.12	.02	.00
Vapor700	.05	.00	.10	.00	.39	.48	.74	.54	.01	.00	.08	.00
Vapor850	.05	.00	.14	.00	.25	.11	.02	.04	.00	.00	.03	.00
Vort mid	.00	.11	.49	.05	.02	.02	.00	.00	.00	.02	.02	.18
Vort low	.02	.27	.32	.02	.06	.04	.01	.66	.00	.18	.05	.06

these smaller groupings. As seen in the table, we compute the KS test for each of the relationships of interest across the entire 154-member (11 ensemble) grouping, compute the same for each of the two-ensemble samples, and find the median p-value for each field from the ten groupings. What we find is that, while there are a few ensembles with nearly zero genesis cases (such as FNLq3_height), which do not contribute to the statistical significance of the environmental effects in general, the picture overall demonstrates an important role for genesis precursors.

Starting with the impact of water vapor on genesis at 1800 UT 24th, we see that at the important layers of 700 and 850 hPa, the statistical significance is quite high that the composites of genesis and nongenesis cases are from different distributions, not only across the 11 ensembles together, but also over most of the two-ensemble groupings. Overall, vapor at 850 hPa appears to have a more significant statistical relationship with genesis than does vapor at 700 hPa, as the median chance that these composites (genesis and nongenesis) are from the same distribution is 0.03 and 0.08, respectively. In terms of the impact of vertical shear on genesis, the relationship is strongest 3 h into the simulation, or 15 h before genesis. What we find is that the corresponding composites at this time are significant to a 5% confidence level in 6 of the 10 ensemble groupings. The median p-value, in this case, is 0.03. We suspect that shear values at 0300 rather than 0000 UTC are seen to be more important to genesis because several ensembles' wind fields are not perturbed directly at 0000 UTC and are only differentiated among the members, 3 h later, by the original perturbations of vapor or θ. Similarly, the wind perturbations have adjusted to the thermal fields by 3 h into the simulation and only now provide an accurate picture of how they might affect genesis at 1800 UTC. Vertical shear at 6 and 9 h into the simulation appears to decline in importance, with the median p-values increasing to .11 and .19, respectively. Even though we are calculating an average

within 500 km of the TC centers, after vortex organization, the storm will produce its own shear, thus complicating the picture substantially.

Compared with the notable genesis impacts of water vapor at 850 hPa and vertical shear at 0300 UTC, we find that the relationship between early vorticity and eventual genesis is just as strong. Notice in Table 11 how for low-level (midlevel) vorticity, the p-value is at or below 0.05 in 5 (8) of the 10 ensemble groupings. Thus, while vorticity in the midlevels, compared with lower-levels, has a slightly more significant statistical relationship with genesis occurrence, it appears (via Figure 63) that the genesis relationship with low-level vorticity is more straightforward. To clarify, we note the strength of the low-level wave in the genesis composite of Figure 65b. That the median p-value (Table 11) for midlevel (low-level) vorticity is .02 (.05) denotes that this field, at both levels, has a statistically significant relationship with genesis outcomes, no matter what the nature of the relationship.

Discussion

In terms of the relationship between environmental parameters at the start of the simulation and genesis likelihood 15-18 h later, we have used ROC curves, composite figures, and the Kolmogorov-Smirnov test to determine that these relationships are meaningful. Individually, 850 hPa water vapor, 200–850 hPa vertical wind shear, and the strength of the 850 hPa pre-existing wave provide a notable amount of genesis predictive power.

Despite the fact that background moisture is quite high in the Tropics in August, we have found that near-TC moisture values play an important role in determining whether a pre-existing disturbance will develop into an organized TC. For example, by solely using

a criteria of 13.8 g/kg of moisture at 850 hPa around the pre-existing disturbance, we were able to successfully predict 78% of the genesis cases which occurred and 87% of the null cases which occurred. The impact of initial water vapor at 700 and 600 hPa on subsequent genesis was slightly less strong, but still significant, nevertheless.

Vertical wind shear is another important parameter related to genesis, which is often treated similarly to water vapor. Essentially, certain maximum values of vertical shear are seen as being necessary for genesis, but not sufficient. In our work, we have found that initial values of 200-850 hPa vertical wind shear, within 500 km of the nascent TC center, have a statistically significant relationship to genesis 15-18 h later. We found that by using a vertical shear criteria of 7.6 m/s criteria at 0300 UTC, we were able to successfully predict 69% of the genesis cases and 70% of the null cases which were analyzed at 1800 UTC.

Similar to water vapor and vertical shear, pre-existing vorticity is often seen as a necessary but not sufficient ingredient for tropical cyclone genesis. For instance, compared to all of the easterly waves that come off of Africa each summer and fall, the number of these which lead to TC formation is quite small. Thus, we were surprised to find that, among our 154 members forecasting Ernesto's environment, the strength of the 850 hPa pre-existing wave differed considerably between composites of genesis and non-genesis 15-18 h later. As seen in Table 11, pre-existing vorticity in both the lower and midlevels often had a statistically significant impact on genesis, although we found that the lower-level wave has more of a clearly positive impact on TC formation (Figure 63)

Despite the large number of resolved and unresolved processes ultimately affecting TC formation, we find that pre-existing environmental conditions play a significant role in the TC genesis process of Hurricane Ernesto. While these results have not been

substantiated across a large sample of storms, we have used genesis/nongenesis composites, ROC curves, and bootstrapped Kolomogorov tests to demonstrate the importance of our varied pregenesis environmental fields on the likelihood of subsequent TC formation. In addition, the variance of pre-existing vorticity, moisture, and vertical shear among our many ensembles and members leads us to believe that these results apply much more generally than for just this case. However, different factors related to TC genesis in the assorted basins, in the different seasons, in various parts of the Atlantic ocean, and for notably different large-scale features may result in notably different statistics than those which were found here.

Highlights of Chapter 7

- Near-TC 700-850 hPa moisture is a strong determining factor of whether a pre-existing disturbance will develop into an organized TC.
- Initial vertical wind shear plays a significant role in determining which members forecast genesis 15-18 h later.
- The strength of the pre-existing lower-level wave has a large impact on subsequent genesis likelihood.
- The relationships between 0300 UTC shear and 0000 UTC water vapor with 1800 UTC genesis are significant to a 5% confidence level for several two-ensemble, or equivalently, 28-member samples.
- The strength of these relationships is unexpected because 1) high moisture values are common in the Tropics in summer and 2) there are many easterly waves which do not become TCs.

CHAPTER 8

STOCHASTIC BACKSCATTER AT HIGH RESOLUTION

Introduction

While the previous chapters (2-6) of this work focused on error growth resulting from differences and uncertainty in model initial conditions, this chapter, like that of our multiphysics ensemble, will again address model error. These two different types of imperfections are acknowledged to generally encompass what is called forecast error. Model error has generally been studied less and the community's knowledge of how to account for it when creating ensembles is notably more limited compared with the uncertainty related to initial conditions (Reynolds et al. 2008).

Whereas initial condition-related uncertainty arises more from a lack of accurate atmospheric observations, model error results from design choices that relate to core aspects of the model, such as spectral, grid point, or finite element; the vertical coordinate system; and the advection scheme and its related truncation error (Shutts 2005). In addition to the varying effects of these choices, parameterizations of subgrid-scale processes also contribute significantly to model error (Teixeira and Reynolds 2008).

Even though the detrimental effects of model error have been difficult to rectify, a few notable categories of solutions to the various sources of this type of error have been developed and examined. First, researchers have used multimodel ensembles in order to overcome the idiosyncrasies associated with one particular model (Krishnamurti et al. 2000; Kharin and Zwiers 2002; Hagedorn et al. 2005). As has been mentioned, when

creating an ensemble, one of the goals is to have the spread among the ensemble members match the error inherent in the forecast. Partially because of model error, however, ensembles are generally underdispersive, which means that the members do not generally account for all the possibilities inherent in a particular forecast (i.e., the truth falls outside the realm of ensemble-predicted possibilities too often) (Berner et al. 2009). By using multiple models in a single ensemble forecast, one is better able to help mitigate the underdispersion through the incorporation of additional uncertainty.

As noted by Fritsch et al. (2000), "variations in model physics and numerics play a substantial role in generating the full spectrum of possible solutions." In terms of accounting for error associated with model physics, certain groups have used different parameterizations within an ensemble prediction system, which we have called a multi-physics ensemble (Houtekamer et al. 1996). Other projects have attempted to reduce the physics-related source of error by developing formulations that impose a stochastic term to the physical parameterizations (Buizza et al. 1999; Teixeira and Reynolds 2008).

Recently, attention has focused on measures that counteract the fact that in NWP models there is an unrealistically large energy sink, which is primarily due to numerical advection error and horizontal diffusion (Shutts 2005). Since subgrid-scale variability is not well resolved in NWP models, the kinetic-energy spectra of the accompanying forecasts drops off much too steeply for wavelengths below 400 km. Because of this and the details of the numerical integration scheme (Berner et al. 2009), the corresponding simulations do not produce the observed $n^{-5/3}$ inertial-range power spectrum found in observations by Nastrom and Gage (1985). Shutts (2005) argued that routine kinetic energy loss is an underlying problem in terms of both numerical integration schemes and parameterizations, and asserted that "a suitably contrived near-grid-scale stochastic forcing function could be used to inject energy back into the model." Before moving into

the details of such a backscatter method, we would be wise to note that this method is a fundamentally different approach to account for model error than that which uses multiple stochastic physics schemes. While stochastic parameterizations, such as in that which is operational at ECMWF (Buizza et al. 1999) or in the Navy Operational Global Atmospheric Prediction System (NOGAPS; Teixeira and Reynolds 2008), attempt to account for the "underlying statistical nature" of the model parameterization scheme, the backscatter methods address a missing model dynamical process, which essentially is the two-way exchange of kinetic energy across the model truncation boundary (Shutts 2005). In other words, instead of "merely sampling subgrid-scale variability by picking realizations from a distribution centered on the value of the deterministic bulk parameterization (Buizza et al. 1999; Lin and Neelin 2002)" the backscatter schemes add "perturbations that mimic the influence of altogether unrepresented subgrid-scale processes" (Berner et al. 2009).

One important component of the backscatter methods is the fact that they perturb the dynamic state directly. Berner et al. (2011) argues that, since the dynamics feed into the physics, this is much better than perturbing the physics tendencies directly (as done in Buizza et al. 1999), which can introduce inconsistencies between the dynamics and physics. These can be especially detrimental, because the model will want to correct such inconsistencies at the next time step, which can produce gravity waves and other spurious artifacts.

The first basic implementations of these methods was accomplished with large-eddy simulations (Mason and Thomson 1992). It was not until more recently, however, that a backscatter method was implemented into a modern NWP model, which was first accomplished by Shutts (2005) with his cellular automation stochastic backscatter scheme (CASBS). His general approach was to use a cellular automaton to generate

evolving patterns, which, together with a dissipation function, ultimately defined a stream-function forcing field. His cellular automaton method largely represented temporal and spatial correlations of the atmospheric mesoscale. Whereas Palmer (1997) envisioned such a pattern depending on available potential energy, Shutts' (2005) cellular automaton (CA) did not depend on any forecast model, and was instead a simple pattern generator. Essentially, this CA pattern, after being scaled by the square root of the dissipation rate, is proportional to the method's stream-function forcing.

It should be mentioned here these authors are not only looking to counter kinetic-energy dissipation due to advection and diffusion, but also that which is due to mountain drag and deep convection through the accompanying parameterization schemes. This latter feature is especially important to note, as these processes were found to count for nearly as much kinetic-energy dissipation as did numerical dissipation in general (Shutts 2005). Countering the total dissipation rate, which includes each of these contributions, with his CASBS method, Shutts (2005) found that the ECMWF model consequently benefited in terms of probabilistic measures of forecast skill and in terms of correcting the previously-too-steep ECMWF spectral slope towards the $k^{-5/3}$ rate typically found in the mesoscales.

Later, a stochastic energy backscatter scheme (SKEB) was implemented and evaluated by those at the Meteorological Service of Canada (MSC; Charron et al. 2010). In this study, the authors provide a general overview of updates made in 2007 to the MSC Global Environmental Multiscale (GEM) model and, at the same time, test the impact of their energy backscatter scheme. In this particular energy backscatter implementation, Charron et al. (2010) injected energy between wave number 40 and 128 using pattern generation methods similar to those of Li et al. (2008). What they found was that, surprisingly, their version of the backscatter scheme caused the GEM model to produce

notably more bias in the low-level temperature field compared to the GEM simulations without the backscatter scheme. The authors admitted that the physical mechanism behind these biases was not clear, but note that "stochastically perturbing a nonlinear system can change its mean state" and cite Palmer (2003). Despite this, however, the dispersion of the simulations using the backscatter scheme was notably higher than the nonbackscatter simulations in terms of temperature and zonal winds at 850 hPa, as well as 500 hPa geopotential heights.

In addition to the general tests of the backscatter scheme as it was used in GEM, Charron et al. (2010) also examined the effect of the SKEBS forcing on the rotational component of the horizontal winds and compared it with a corresponding forcing on the divergent component of the same winds. Comparing both methods against the GEM version without SKEB, they found that forcing the divergent winds had a small impact in terms of increasing the 500 hPa geopotential height power spectra compared to the SKEBS which forced the rotational wind components. The difference notably decreased with time, and Charron et al. (2010) attributed these results to the "well-known" principal that rotational modes are more likely to inverse cascade than divergent modes. Overall, they found that SKEB, in terms of a forcing of the rotational wind components, helped improve forecast reliability by primarily improving ensemble dispersion.

Berner et al. (2009; 2011) have largely been responsible for implementing and evaluating more recent and widely-used versions of the backscatter method. In their 2009 study, instead of using a cellular automaton as did Shutts (2005), they used a first-order autoregressive process on each spherical harmonic of the streamfunction forcing. They did this such that they would have control not only over spatial and temporal correlations, but also in terms of the spectral characteristics of the perturbations. In addition, cloud-resolving models were used to inform parameters of the backscatter scheme in terms of

the "power-law exponent of the forcing streamfunction" (Berner et al. 2009). By implementing such a method in the ECMWF ensemble prediction system, the same authors found that they achieved a better spread-error relationship, improved rainfall forecasts, and better probabilistic skill compared to simulations without the backscatter scheme. While Berner et al. (2009) found improvement using a simple backscatter based on a constant dissipation rate, the best results came from "flow-dependent formulations of the unresolved processes."

More recently, modified backscatter schemes have been implemented in the Met Office's Global and Regional Prediction System (MOGREPS; Bowler et al. 2009; Tennant et al. 2011). The same studies report a positive impact on forecast reliability and probabilistic skill from their backscatter scheme implementations. While the SKEBS scheme was first implemented in larger-scale ensemble systems, Berner et al. (2011) adapted the method to the Weather Research and Forecasting (WRF) model. One of the main adaptations they had to make when converting the backscatter scheme from the pseudospectral-core global ECMWF system, to that of the WRF, which is a limited-area model that uses finite differences, was to change the basis functions of the stochastic kinetic-energy backscatter "from spherical harmonics to 2D-Fourier modes" (Berner et al. 2011). The same authors compared this scheme to both an ensemble based on 1) various physics parameterizations and 2) a combination of the backscatter scheme combined with the multiphysics and found that the stochastic backscatter scheme outperformed the ensemble using multiple combinations of different physics schemes. In general, however, they found that the best performing ensemble was that which combined the multiphysics scheme with the stochastic energy backscatter scheme.

Overall, the stochastic kinetic-energy backscatter schemes usually cause their ensembles to produce better spread-error relationships and altogether help solve the

general ensemble underdispersiveness problem. In addition, these schemes alleviate the unnaturally steep spectral slope nearly all ensembles possess in the mesoscale and help them move closer to the $k^{-5/3}$ slope found in nature. At the same time, this works to decrease error at the small scales which subsequently lessens the up-scale error propagation that consistently leads to general forecast error.

While the SKEBS scheme has been shown to generally ameliorate the ensemble underdispersiveness problem to some extent, there are a number of interesting questions and issues that have arisen from the results of past research. First, much of the research into the various SKEBS implementations has examined ensemble resolution at or above 45 km (Shutts 2005; Charron et al. 2010; Berner et al. 2011). Even with the help of SKEBS, these relatively-low resolution ensembles, along with one at a higher 25 km resolution (T799; Shutts 2005), were not able to effectively replicate the $k^{-5/3}$ spectral slope. In addition, little work has been done to examine the impact of the SKEBS scheme on modern mesoscale forecast models. For example, as described early in the dissertation, the author's group has successfully implemented ensemble generation capabilities of 14 members at 5 km resolution. While this higher resolution, and thus lessened dependence on physics parameterizations, somewhat inhibits the kinetic-energy spectral drop off at small scales, these mesoscale models suffer the same energy-dissipation issues as the global models in operation at the various national centers. Despite this, however, relatively few, if any, studies have examined the resultant spectral effect of the SKEBS scheme's injection of energy at the small scales in such ensemble systems. Also, while previous studies have demonstrated that the SKEBS forcing on the rotational wind has a larger effect than that of the divergent wind, these results were also based on models whose resolution was much higher than our 5 km. While Charron et al. (2010) said that it was "well-known" that rotational modes are more likely to inverse

cascade than divergent modes, the effect of this on SKEBS implementation has not been examined in a high-resolution mesoscale ensemble.

In addition to this uncertainty regarding the resultant spectral effects of SKEBS implementations in high-resolution models, the ability of the scheme to increase dispersiveness has only thus far been lightly touched upon in the relevant research. This elevated dispersiveness was indeed seen as being the most important component of SKEB's overall beneficial impact on forecast reliability by Charron et al. (2010), but, as Berner et al. (2009) mentioned, errors have to remain the same or decrease at the same time for forecast reliability to increase. One of the more complicated SKEB-related results examined thus far is the fact that Charron et al. (2010) found that the scheme degraded low-level temperature bias compared with their non-SKEBS experimental forecast. Such behavior certainly will not help reduce the ensemble system's reliability in terms of that particular field unless the increase in dispersiveness is truly remarkable.

Thus, there are large gaps in the community's knowledge of what kind of forecast benefit popular SKEBS schemes provide. This is partially in terms of whether vapor, temperature, or geopotential heights increase in dispersiveness and/or decrease in terms of bias. A related question is whether the spread between these values is primarily improved, compared against a regional control ensemble, more towards the boundary layer or in the middle troposphere. In addition, as the kinetic-energy is injected at small-scales the particular wavelengths at which these error/dispersion relationships improve is of great interest. Also important is characterizing the rate of upscale error growth due to unresolved processes in SKEBS versus non-SKEBS forecasts. Determining both where progress is being made and where SKEBS is actually undermining our efforts in terms of the spread/error relationship is an important factor in determining whether the scheme is worth the computational cost. Relatedly, these comparisons against not only a GEFS-

based control ensemble, but also in terms of forecasts using the various SKEBS parameters available will also benefit the community because of the general lack of data on how to best tune the scheme.

Another large consideration when studying the SKEBS scheme is the region where one is making forecasts. As briefly mentioned, not only do diffusion and advection cause kinetic-energy dissipation in the models, but so do mountain drag and deep convection as well. As explained by Shutts (2005), this arises because most convective parameterization schemes do not explicitly account for the fate of the kinetic-energy released by buoyancy and instead focus on the thermodynamic impact of the subgrid-scale clouds. Thus, the various SKEBS designs implement their procedures using a total dissipation rate, which includes contributions from deep convection, along with diffusion and advection. Since deep convection is obviously most concentrated in the Tropics, the dissipation contribution from this component will be largest in that region of the globe and, consequently, the positive effects of SKEBS can be much more pronounced in that region. This was seen in Berner et al. (2009), who found that the difference between spread and error in their ECMWF simulations was reduced, due to their SKEBS implementation, in the tropical regions compared to the midlatitudes. This was not only seen in their spread/error relationships, but also in terms of Brier skill scores and ranked probability skill score. Due to these practical considerations in terms of SKEB's actual construction and the corresponding results from Berner et al. (2009), it is apparent that not only are particular geographical considerations important when assessing SKEB, but also that the best place to examine subtle benefits of SKEBS vs non-SKEBS forecasts is likely in the Tropics.

Due to the uncertainties related to the effect of the various kinetic-energy backscatter schemes and the other practical details of their construction and

implementation, we determined that an analysis of carefully constructed SKEBS schemes at high resolution as compared to the ensembles treated in the first chapter of this dissertation would be beneficial and enlightening on several fronts. Not only is our ensemble prediction system setup at 5-km resolution able to create similarly designed, in terms of timing and domain, 14-member SKEBS forecasts, but we can easily take advantage of our continuing tropical focus to determine any subtle SKEBS forecast characteristics in terms of differences from our prior GEFS-based control ensemble and our varying bred ensembles. In addition to these natural advantages that arise from comparing carefully designed control, breeding, and SKEB-based ensembles, another important consideration is the relative impact on TC genesis. Considering the fact that these backscatter schemes inject energy that was lost due to subgrid-scale processes, the various ensembles' portrayal of genesis might yield interesting physical differences, as the community's lack of TC forecast accuracy often comes due to their inability to resolve genesis-relevant processes on the small scale (Hennon and Hobgood 2003).

In light of these practical and theoretical considerations, the latter part of this dissertation will use SKEB-based 5-km resolution ensembles of Hurricane Ernesto in order to better determine the effect of the backscatter algorithm on forecast error characteristics, error spatial spectral slopes, and TC genesis likelihood. These high-resolution SKEB-based simulations will be examined in light of our previous CNTL simulation of Ernesto based on the GEFS and also as compared to the varying breeding ensembles forecasting the same event using strategically different implementations. Specific questions to be answered in this chapter include the following: 1) does the SKEBS ensemble show statistically significant spatial spectral differences as compared with a similar-resolution GEFS-based ensemble; 2) does the backscatter ensemble show an energy slope of $k^{-5/3}$ below 400 km wavelength more so than the BGM and control

ensembles; 3) does the ensemble bias respond negatively when implementing the SKEBS scheme as compared to the control ensemble; 4) does the increase in dispersion, for key variables, increase more in the SKEBS scheme, compared to the control, than does the ensemble error; 5) does the rotational component of the wind demonstrate a large response to the SKEBS implementation than does the divergent component of the wind; and 6) what, if any, are the key physical differences in terms of TC genesis when the model is being perturbed via the SKEBS scheme as compared to the GEFS's Ensemble Transform scheme?

Ensemble creation and experimental details

In order to achieve our goals and answer the above-stated questions, we use the Weather Research and Forecasting (WRF) model version 3.4.1 and the stochastic kinetic-energy backscatter scheme (called SKEBS in WRF parlance) version 1.0, which is based on the work of Berner et al. (2011). It should be noted that the SKEBS version used here has a flow-*independent* dissipation rate. Berner et al. (2009) found that such a constant dissipation rate can still notably improve the bias and dispersion of an ensemble, and, because it avoids calculating the full dissipation rate, renders the SKEBS implementation simpler and less computationally expensive.

In order to keep the spatial spectrum and dispersion/error comparisons as clean as possible, for these simulations, we keep the same physics parameterizations, initial and boundary conditions, and domain positioning as set in our original regional (control) simulation of Hurricane Ernesto. Since there is no need for a prior breeding period, these SKEBS simulations will start at the same time as the control ensemble—that is, at 0000 UTC 24th Aug 2006, but will finish at 1800 UTC 24th Aug. This simulation end time was moved up 18 h compared to our other ensembles in consideration of 1) the SKEBS'

greater computational expense (up to 3x regular WRF simulations on certain clusters) and 2) the fact that our spectral comparisons and any effects on genesis occur before this new end time.

We first begin by testing the WRF SKEBS solely on one domain at 45-km resolution in order to properly test and calibrate the scheme for our particular forecast. In terms of the scheme's implementation, there are several parameters one can tune in order to achieve a particular calculation. For example, while the dissipation rate is constant, one can alter the amplitude of both the rotational wind perturbations (psi, or stream function) and the potential temperature perturbations. In addition, one can also alter the vertical structure of the random pattern generator between "random phase" and a more "barotropic" calculation. Overall, in these experiments, we primarily use a standard SKEBS configuration to test the backscatter result against our previous ensembles. In addition, however, we create several SKEBS-based ensembles in order to determine the importance of its tuning parameters, and to better determine the significance of any notable spatial spectra differences from our previous 5-km resolution ensembles. It is thought that the SKEBS scheme may have particularly interesting physical effects on TC genesis, as a major part of its implementation is effected through rotational wind perturbations. Thus, we will analyze the results of the variously tuned SKEBS schemes in terms of their potential impacts on TC genesis processes.

After preliminary experimentation of the backscatter scheme with various parameters at 45-km resolution, we then implement the same 5-km inner domain and double-nested structure seen in our earlier ensembles (Table 1). After continued testing with these nested domain specifications, we found that in order to achieve somewhat realistic error spectra results, SKEBS must be activated for not only the outermost of the three domains, but also on each of the nested domains as well. Having established this,

we then carefully alter the rotational wind, potential temperature, and vertical structure parameters to create three unique 5-km resolution ensembles. First, having established with our 45 km resolution simulations that the potential temperature perturbation produces much more realistic TC intensity and intensity spread with a fixed amplitude of 2.0E-6 instead of 1.0E-6, we use the former as our basis and can then turn toward the other parameters. Thus, the first 5 km resolution SKEBS ensemble of note is created with this 2.0E-6 amplitude θ perturbation, a standard 1.0E-5 amplitude rotational wind perturbation, and a constant vertical pattern generator (Table 12) and is called the backscatter control, or BS_CNTL. Keeping everything else the same as ensemble one, the second ensemble alters the vertical pattern generator such that it has a random phase (called BS_VERT) and, keeping all else the same as in BS_CNTL, the third ensemble changes the rotational wind perturbation amplitude such that it is doubled to 2.0E-5 (called BS_PSIDOUBLE; Table 12).

Bias over time

As a preliminary examination of the water vapor and θ bias of the SKEBS ensembles, we again refer to Figure 8. In general, we find that the SKEBS θ bias is notably higher than most of the initial condition-based ensembles. This is especially exacerbated between 800 and 900 hPa and, to a lesser extent, 400 to 800 hPa. What this means is that the three SKEBS ensembles are producing θ values, domain wide, that are higher than those of the rest of the ensembles. In terms of the differences between the backscatter methods, it appears that the random vertical pattern generator in BS_VERT mitigates the heightened θ bias somewhat. Interestingly, in terms of vapor, the SKEBS ensembles produce notably lower absolute values of bias, over this time period, at almost

Table 12

Configuration of SKEBS ensembles and their parameters

Ensemble experiments	Th Amplitude	Psi Amplitude	Vertically random
BS_CNTL	2E-6	1E-5	Off
BS_VERT	2E-6	1E-5	On
BS_PSIDOUBLE	2E-6	2E-5	Off
BS_CNTL_NURI	2E-6	1E-5	Off

all heights. The improvement of the SKEBS ensembles compared to the others here is quite remarkable, especially considering the large amount of averaging that takes place in the calculation. This helps demonstrate that perturbation methods taking into account model error can produce fairly consistent and substantial differences from initial condition-based ensembles even within one TC forecast case. Overall, since the bias differences among the SKEBS ensembles vary notably by height, we will refrain from drawing any conclusions with regards to the particulars of the SKEBS implementations here. We will simply say that our WRF-based implementation of SKEBS produces elevated θ bias, especially below 400 hPa, which confirms what was found by Charron et al. (2010) in their SKEBS experiments. At the same time, our SKEBS ensembles markedly reduced water vapor bias compared with our initial condition-based ensembles.

Spectral characteristics

Next, we examine these ensembles in terms of their spectral patterns by comparing them with several of our previous ensembles and amongst the SKEBS simulations themselves. Figure 68 presents the results of a Fast-Fourier Transform calculation for the CNTL, FNLq3, FNLth3, and BS_CNTL ensembles. Similar to our previous spectral figures, this calculation is performed in terms of meridional patterns, which result is then averaged zonally and over the ensemble. Unlike the previous calculations, this one is done for *both* ensemble error and perturbation size. Here we have chosen to average from twelve to 18 h into the simulation to allow time for the SKEBS scheme to manifest itself in the data as well as discard any short-term artifacts.

First, notice the notably larger perturbations produced by the BS_CNTL ensemble compared to the others, over much of the spectrum. In terms of the spectral power at

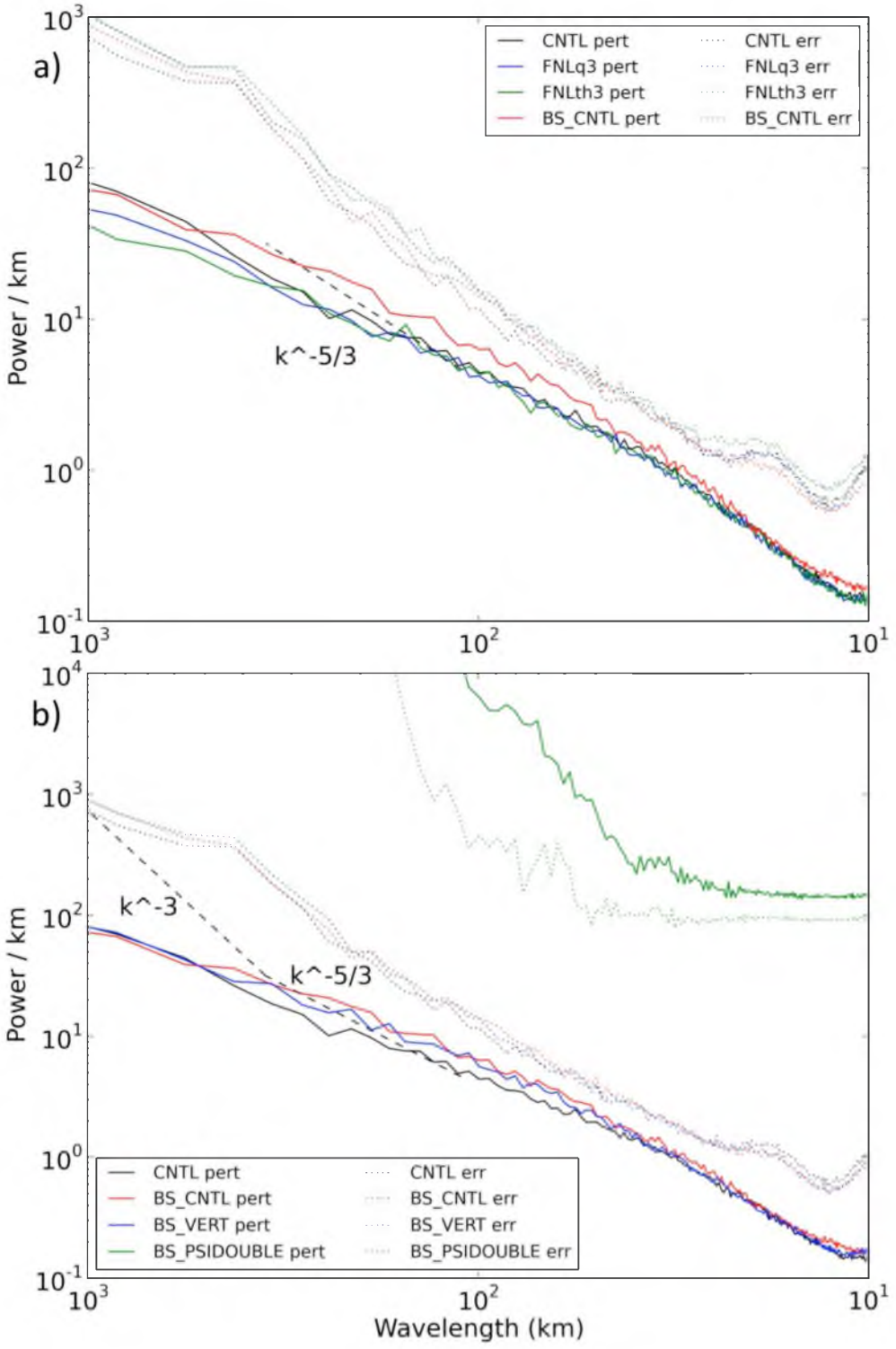

Figure 68. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height member differences from analysis (errors) and from ensemble mean (perturbations), calculated across the domain, then across a zonal and ensemble average from 1200 to 1800 UTC 24th for a) various types of ensembles and b) different SKEBS implementations.

small wavelengths, we notice that, below 100 km wavelengths, the backscatter control ensemble (BS_CNTL) displays only slightly higher perturbation values than the other ensembles (Figure 68a) compared with its advantage at 100 km, thus giving it a steeper slope in this region. Because of how it performs at wavelengths greater than 100 km, however, the slope from 400 to 100 km is shallower for the BS_CNTL ensemble than it is for the breeding and control ensembles. Thus, any prima facie decrease in spectral slope of the SKEBS, which was one of the motivations for the SKEBS implementation, comes not at the high-resolution truncation boundary, but from roughly 400 to 100 km.

Note that we find that for all ensembles (in Figure 68a) the spread, or perturbations, are consistently smaller than the ensemble error. This occurs for all wavelengths and throughout the simulation (not shown). We see that from 1200 to 1800 UTC, the BS_CNTL ensemble is providing larger perturbations than are the control and breeding ensembles from 20 to 300 km wavelengths, without showing notably more error (Figure 68a). Thus, we find that compared with our previous ensembles, the BS_CNTL ensemble provides ensemble perturbations that more closely match forecast error (i.e., less underdispersion), and that this difference is centered around the 100 km wavelength. In terms of the significance of these differences, we find them to be quite high, even for similar-looking profiles. For example, using the two-sample Kolmogorov-Smirnov test, we find that the p-value for the differences between the CNTL and BS_CNTL spectral error in this figure is .04. This means that there is only a 4% chance that they are from the same distribution, despite how similar of slope they have.

Turning toward the differences between the various SKEBS implementations (Figure 68b), we find that, besides the anomalous output from the BS_PSIDOUBLE ensemble, there is very little spectral change when creating the ensemble using a random (BS_VERT) or constant (BS_CNTL) vertical pattern generator (Table 12). Regarding the

BS_PSIDOUBLE ensemble, we find that the results are calculated accurately, and that a psi amplitude perturbation specification of 2E-5 appears to provide an excessive streamfunction energy boost, which causes the ensemble to not only exhibit too much spread, but excessive error as well.

The $k^{-5/3}$ slope which has been found in observational data (Nastrom and Gage 1985) but has been difficult to produce in model simulations, is fairly consistently found at wavelengths smaller than 100 km in our ensembles. Such a $k^{-5/3}$ slope was typically found to extend out to around 400 km wavelength (Hohenegger and Schar 2007; Rotunno and Snyder 2008), but in our ensembles, the slope is notably flatter than $k^{-5/3}$ from 100 to 400 km and, indeed, does not show any sign of the k^{-3} slope at wavelengths greater than 400 km. This is similar to what was earlier found with the BGM and CNTL ensembles. Similarly, these SKEBS differences with the observations of Nastrom and Gage (1985) at the large end of the spectrum are likely due to the fact that a regional model is being used here and, thus, we accept slope inaccuracies at this end of the spectrum to get a high-resolution view of SKEBS' effect on the small wavelength part of the spectrum.

Because of the fact that the ensemble creation schemes are generally designed for midlatitude baroclinicity (Snyder et al. 2010), we are interested in evaluating how our breeding, control, and SKEBS ensembles perform both in the Tropics in general and close to Ernesto's center, which provides a contrast between the more barotropic error growth happening at large radii and the more baroclinic and convective-style error growth more prevalent near the TC.

In consideration of this, we calculate the FFT-based spectral slope of the same ensembles in Figure 68, but we do it over a 1000 km-per-side box centered on simulated Hurricane Ernesto's center at that time. For simplicity, we will call this the 1000 km proximity area. The calculation remains one of the meridional errors and perturbations, as

averaged in the zonal direction within the box, and over the ensemble. The result is plotted in Figure 69. What we find is that except for a drop off in energy above 300 km due to the smaller area included in this calculation, the general shapes are similar compared to the domain-averaged figures in Figure 68. Nevertheless, we find that the there are notable differences in the details. For example, compared with the calculation across the domain, here we see that the differences between error and perturbation size are notably smaller (Figure 69a). This decrease in underdispersion is primarily based around 90 km wavelength, but is also present up to 200 km and as low as 20 km. It even appears that for the BS_CNTL scheme (Figure 69a), the error and perturbation sizes are nominally the same around 80 km wavelength. While the error/spread ratio does improve slightly when moving from the entire domain to the 1000 km proximity area for the bred ensembles also, this difference is smaller than that for the control and BS_CNTL schemes. Considering their initial positions (Figure 68), this leaves the SKEBS and control ensembles with a notably more favorable spread and error profile compared to the breeding ensembles, in the 1000 km proximity area. Note that these improvements in error/spread ratio near the TC come largely from a decrease in error compared with values far from the TC.

In terms of the differences between the SKEBS schemes' performance within this 1000 km proximity area (Figure 69b), we find that the BS_CNTL and BS_VERT ensembles behave quite similarly, with error/spread ratios similar to that of the control ensemble near 100 km. Below this wavelength, however, the SKEBS schemes show remarkably smaller error than in the CNTL ensemble. The BS_PSIDOUBLE ensemble continues to show spectral energy that is much too high, and is included for the sake of completeness.

Wanting to isolate the TC-proximate effects of convective error growth even

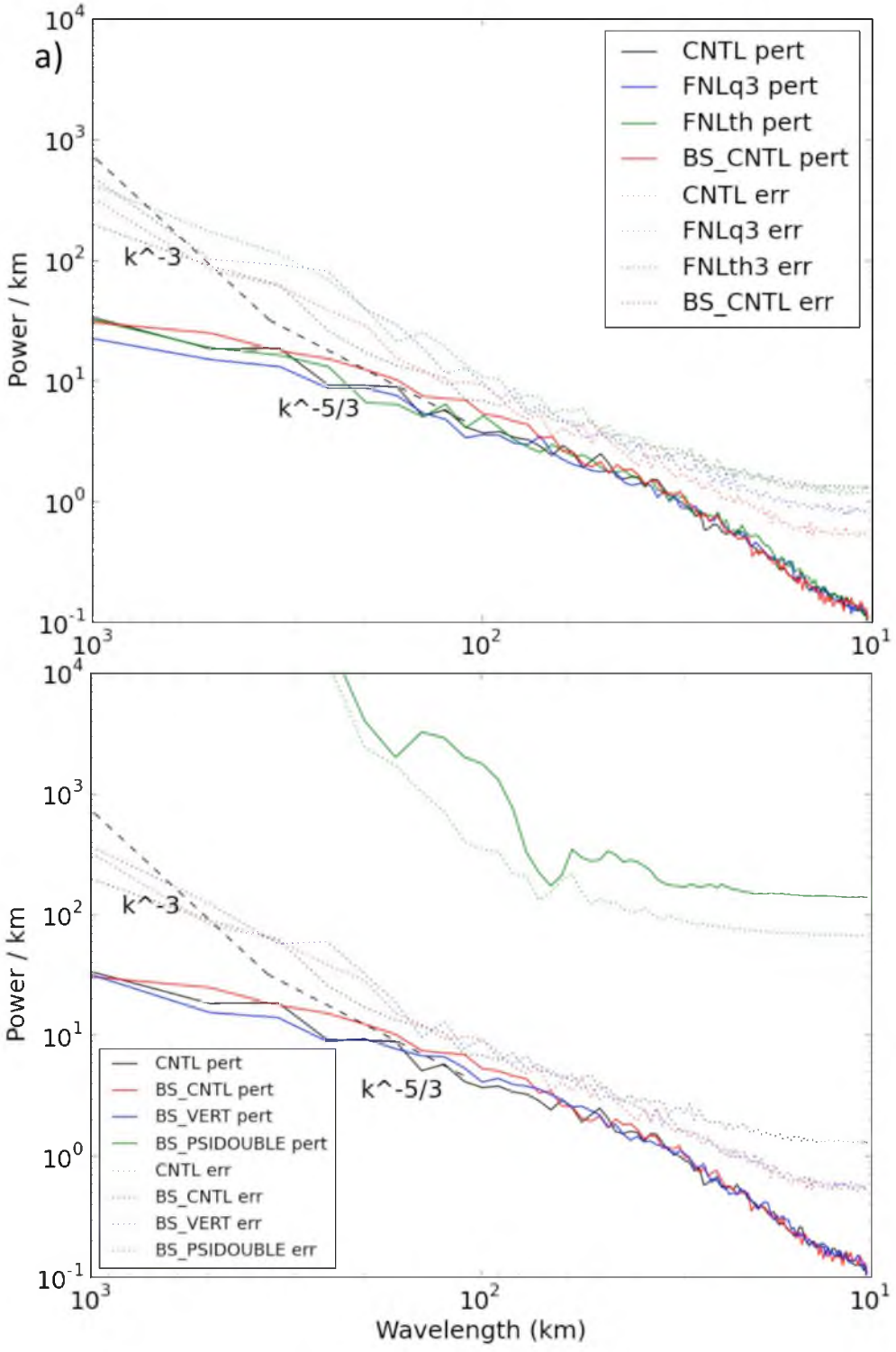

Figure 69. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height member differences from analysis (errors) and from ensemble mean (perturbations), calculated in the meridional direction, then across a zonal and ensemble average, within a 1000-km-per-side box, centered on the simulated TC, from 1200 to 1800 UTC 24th for a) various types of ensembles and b) different SKEBS implementations.

further, we calculate the FFT-based error and perturbation spectrum for the same ensembles within a 200 km-per-side box. This region is called the 200 km proximity area and the resulting calculations are plotted in Figure 70. What we find is that in this smaller region, the breeding ensembles have a lower error/spread ratio than does the CNTL ensemble, which is a reversal from our larger domain spectral calculations. This is primarily due to the fact that the ET-based CNTL ensemble shows notably higher error than here than do the other ensembles. Despite these points, however, the BS_CNTL ensemble, and the SKEBS ensembles in general, show the lowest error/spread ratio among our various categories of perturbation-creation methods, not only in these TC proximity areas, but also across the domain in general.

Considering that Charron et al. (2010) found that forcing the rotational modes, compared with divergent modes, in their SKEBS scheme produced a power spectra closer to that $k^{-5/3}$ slope found in nature, we wanted to examine these separate components in our simulations. To do this, we split the wind into rotational and divergent components and calculated the spatial power spectra over the last 6 h of the simulation period. This is performed for the various SKEBS-based ensembles as well as the CNTL ensemble. The result, plotted in Figure 71, provides a way to determine whether our WRF-based SKEBS implementation provides more energy on the small scales, compared to the CNTL case, in terms of rotational or divergent wind components.

What we find in Figure 71 is that the divergent and rotational components of the wind do not exhibit significant differences from each other, no matter the ensemble considered or whether one examines ensemble error or perturbation size. In terms of comparing the CNTL ensemble against the BS_CNTL ensemble in terms of the rotational or divergent components, we find that the SKEBS-based ensemble does not provide any

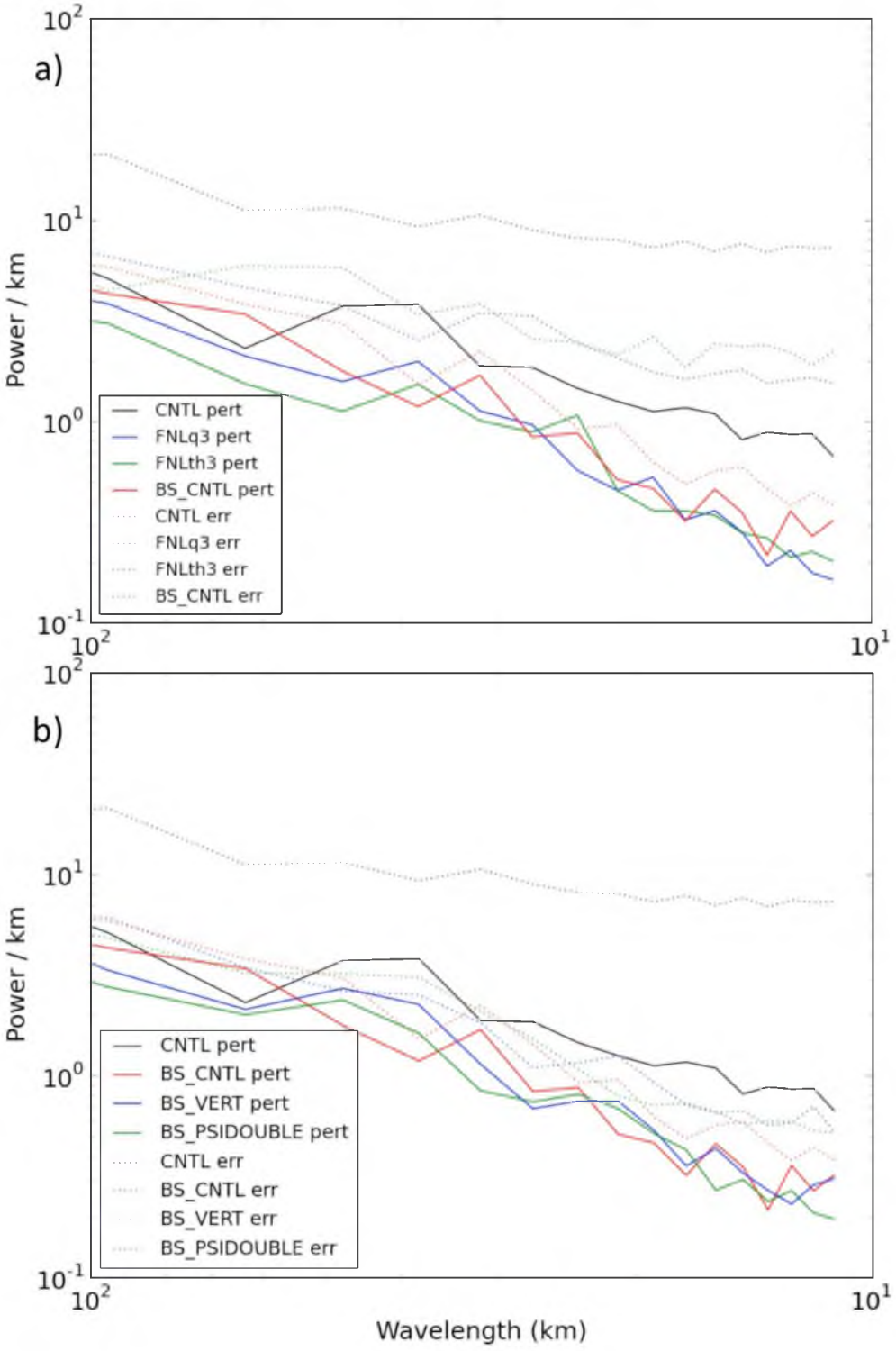

Figure 70. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height member differences from analysis (errors) and from ensemble mean (perturbations), calculated in the meridional direction, then across a zonal and ensemble average, within a 200-km-per-side box, centered on the simulated TC, from 1200 to 1800 UTC 24th for a) various types of ensembles and b) different SKEBS implementations.

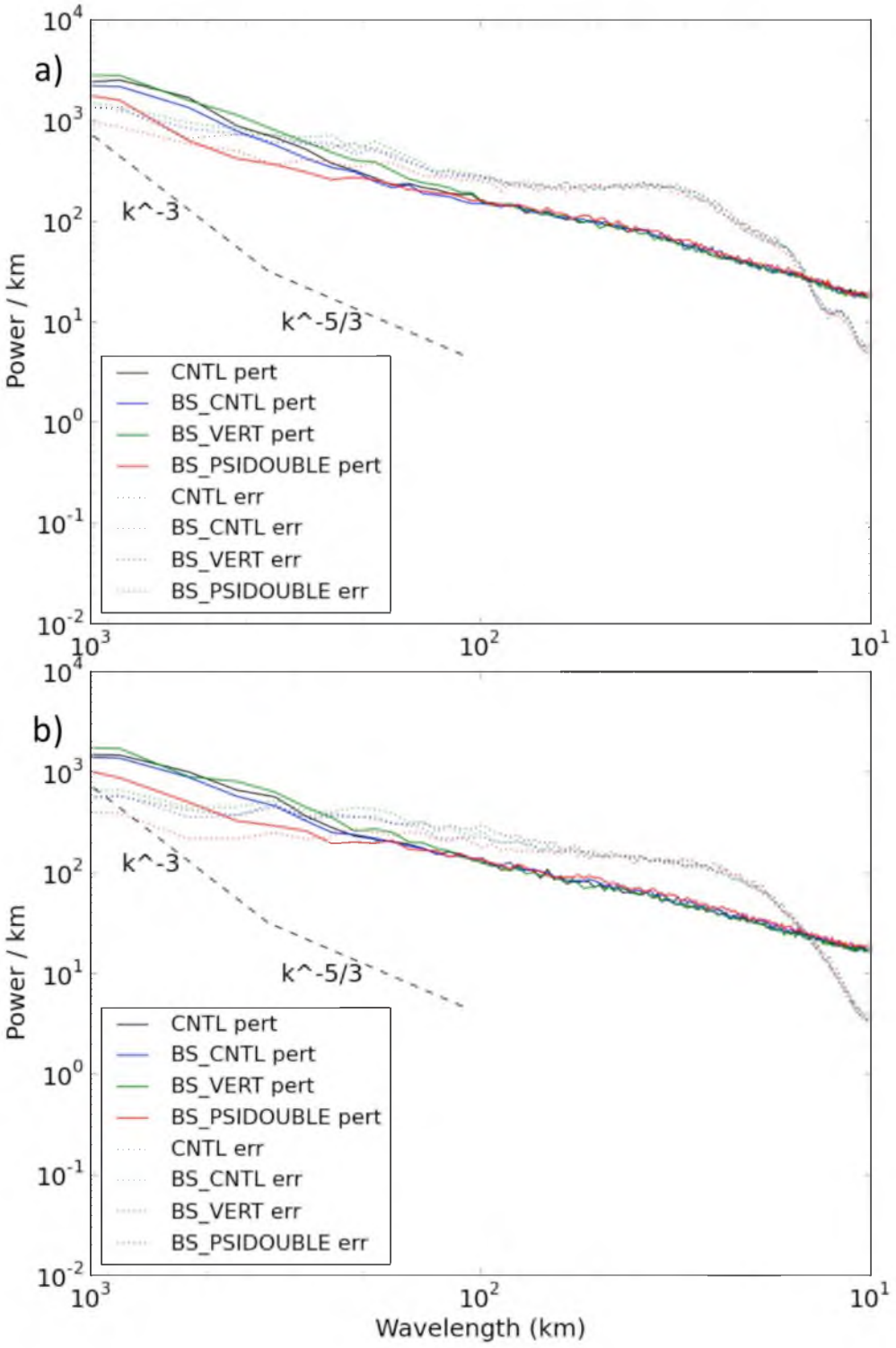

Figure 71. Fast Fourier Transform-based spectral analysis of 500 hPa geopotential height member differences from analysis (errors) and from ensemble mean (perturbations) calculated in the meridional direction, then across a zonal and ensemble average, from 1200 to 1800 UTC 24th, for a) the rotational and b) the divergent component of the wind.

significant boost in perturbation energy at any points in the spectrum. While the BS_PSIDOUBLE ensemble shows some spurious effects in terms of geopotential heights (Figure 68), we note that in terms of wind, its behavior is somewhat in line with that of the other SKEBS ensembles (Figure 71). Despite this, however, we do note that by doubling the amplitude of the streamfunction perturbation (i.e., in BS_PSIDOUBLE), we find that the error and perturbation size are significantly lower than that of the CNTL and other SKEBS ensembles for wavelengths greater than 200 km. We see that even though the SKEBS calculations are made in the rotational component of the wind, due to the model integrations, the impact of the backscatter is felt nearly equally in both wind components. Overall, in terms of the wind spectrum, we emphasize the significant impact of the streamfunction doubling, which leads to larger differences from BS_CNTL than the differences between CNTL and BS_CNTL. This, as mentioned by Charron et al. (2010), is likely related to the fact that rotational modes, which are represented by the streamfunction perturbation amplitude doubling in BS_PSIDOUBLE, are quite prone to the inverse cascade effects that can quickly translate the energy backscatter to the rest of the simulation.

Dimensionality of SKEBS spread and error

While we have seen that the error and spread of the SKEBS ensembles were often closer to each other, in terms of spatial spectra, than were those of the CNTL or breeding ensembles, we want to examine how these ensembles compare in terms of other measures of error and spread. We now turn to the E-dimension to determine the way in which the dimensionality of the SKEBS ensembles' error and spread compares with that of the previous, initial condition-based ensembles. As with the previous E-dimension figures, for each ensemble, this metric is calculated over time and averaged over various distances

from the corresponding TC center. Presented in Figure 72 is the domain-averaged E-dimension calculation for a) various ensemble creation methods and b) SKEBS-based ensembles.

What we first notice in Figure 72 is the fact that both spread and error dimensionality have the same general pattern over time, in that they increase sharply for the first 9 h and then plateau afterwards. A similar, early rapid growth rate in dimensionality is seen in both the SKEBS, CNTL, and breeding ensembles. We find that in this domain-averaged calculation, the dimensionality of the perturbations is quite close to, and often higher than, the dimensionality of the ensemble error. This latter phenomenon is most evident in the breeding schemes, and never occurs in our ensembles when considering regular error and spread spectral differences. It thus appears that relative changes in error and spread dimensionality in our various ensembles do not relate well to relative *actual* changes in the relationship between error and spread.

In order to determine how these dimensional spread/error relationships change in the vicinity of simulated Hurricane Ernesto, we now only use those data which fall within 200 km radius of the corresponding TC center. This result plotted in Figure 73 and provides several notable differences compared with the domain-averaged results found in Figure 72. First, note how much faster the dimensionality growth is which occurs within 200 km radius of the TC center, over the first 9 h of the simulation, compared to that which occurs at large radii. For example, we see that by 6 h into the simulation, the near-TC E-dimension values are above 9, whereas when considering points far from the TC, values at this time are near 7. This change is present for all of our disparate kinds

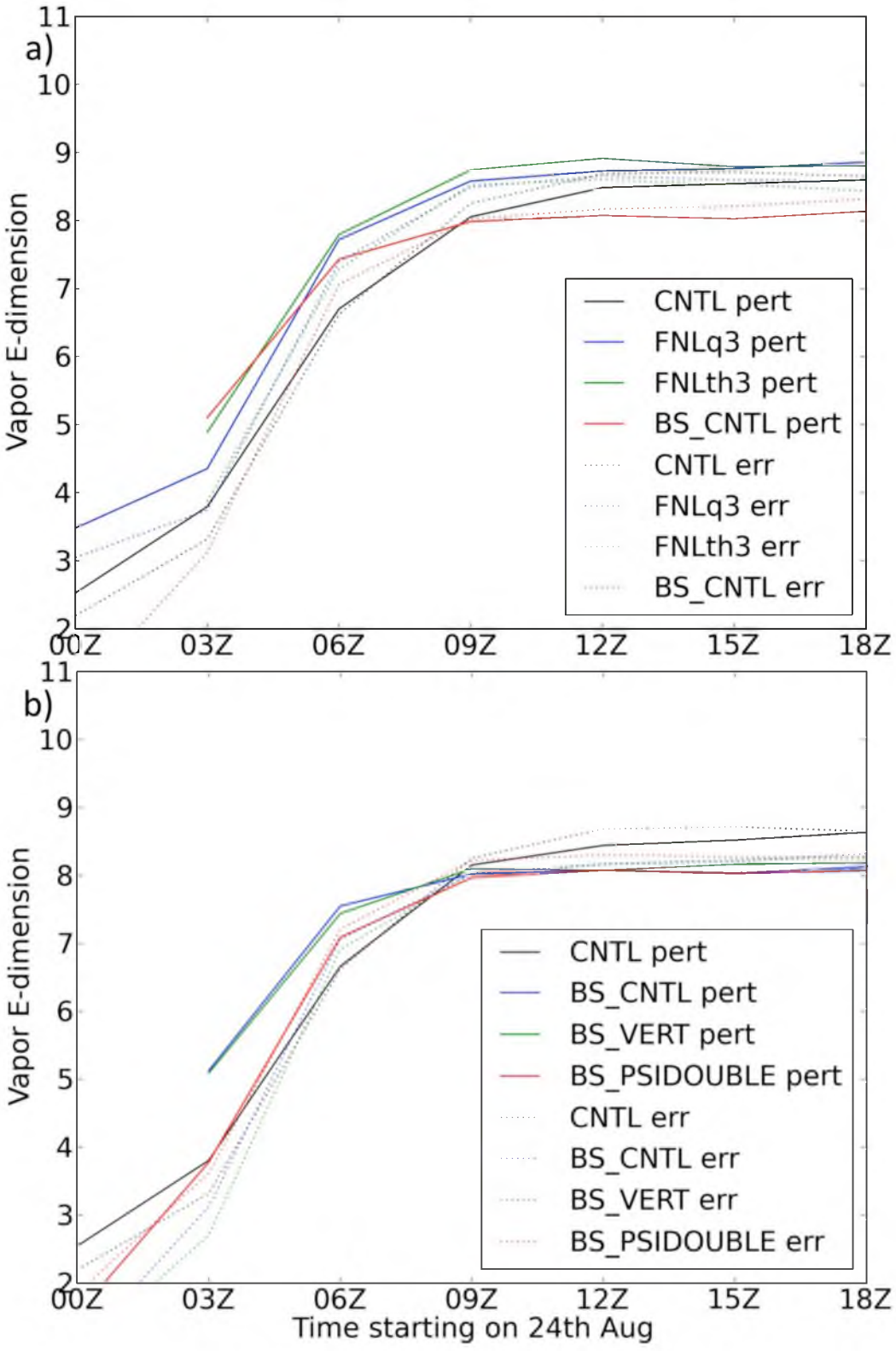

Figure 72. Time series of error and perturbation E-dimension for 850 hPa domain-averaged water vapor for a) various types of ensembles and b) for the SKEBS implementations.

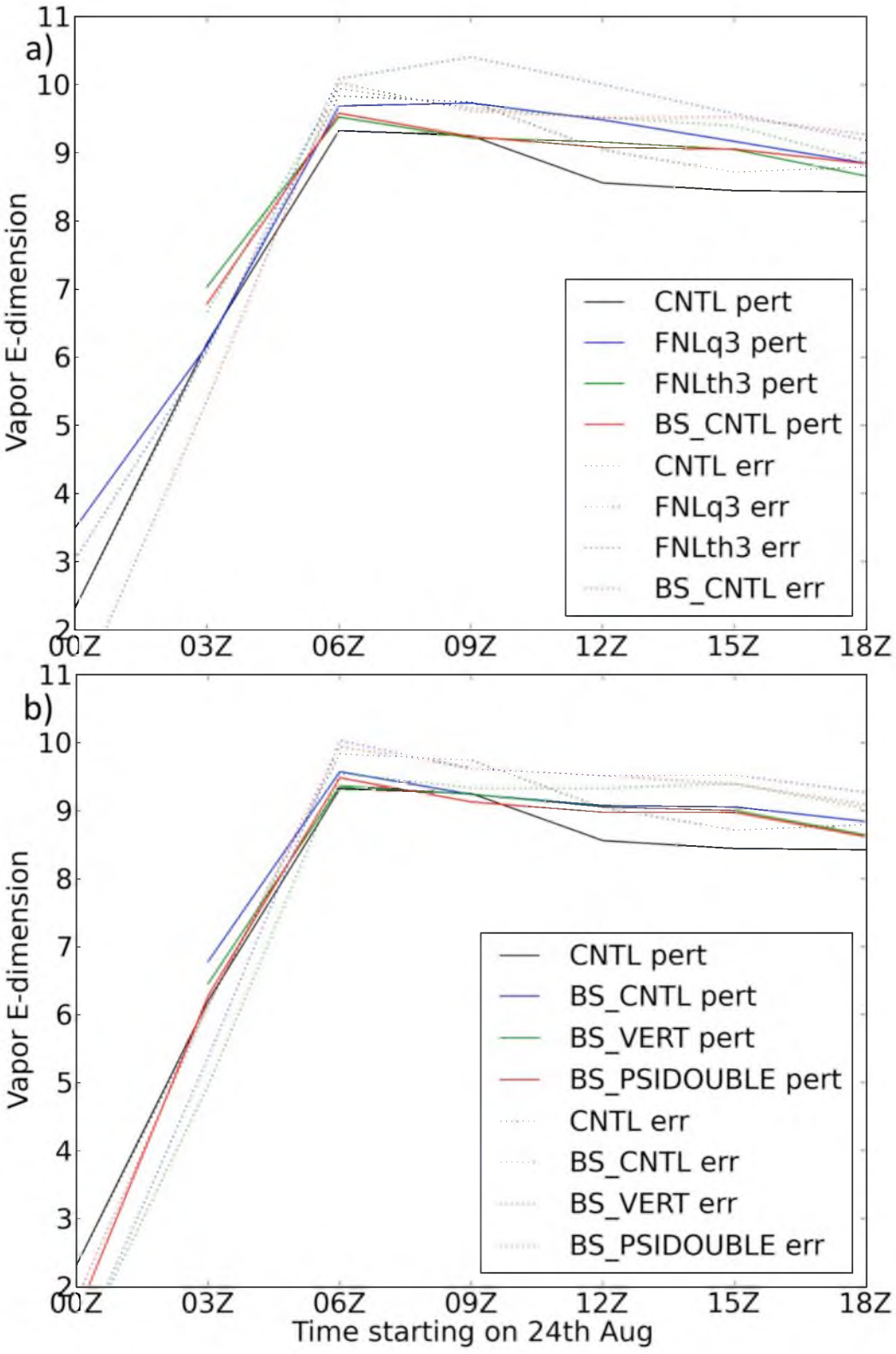

Figure 73. Time series of error and perturbation E-dimension for 850 hPa water vapor over time and within 200 km radius of simulated TC center for a) various ensemble categories and b) for the SKEBS ensembles.

of ensemble creation methods and for both error and spread. In terms of the dimensionality of spread versus error near the TC, we find that the former is consistently lower than the latter, which was *not* seen when considering domain-averaged values in Figure 72. This holds for both breeding, ET (or CNTL), and SKEBS-based ensembles. Not only is the near-TC error dimensionality higher than that of the spread more often at small versus large radii from the TC, but it also generally higher by a greater amount. Thus, overall it appears that TC-related processes tend to increase the number of directions of ensemble variance from the analysis disproportionately compared to the number of directions of ensemble variance from the 14-member mean.

Genesis results

In order to view the physical outcome of the various SKEBS-based methods we have used, we plot a synoptic view of the 14 members in each ensemble at genesis time, which is 1800 UTC 24th August 2006. This is done to not only compare how the near-TC genesis region responds to the varying parameters of the backscatter methods, but also to examine the overall ability of the SKEBS ensembles to spin up TC vortices. Plotted in Figure 74 are the corresponding results of the BS_CNTL ensemble in terms of 850 hPa geopotential height and vorticity. What we find is that there is a significant amount of variability between the members in terms of TC strength and positioning at this time. This variability is somewhat similar to that seen in the CNTL simulations (Figure 9). In the BS_CNTL ensemble, however, the TCs are notably less intense; this is seen in the fact that CNTL shows 7/14 ensemble members whose TC has reached genesis, whereas the BS_CNTL ensemble is 4/14 at this point (Table 13). Also notice the geographically extended geopotential height anomalies throughout the domain in member

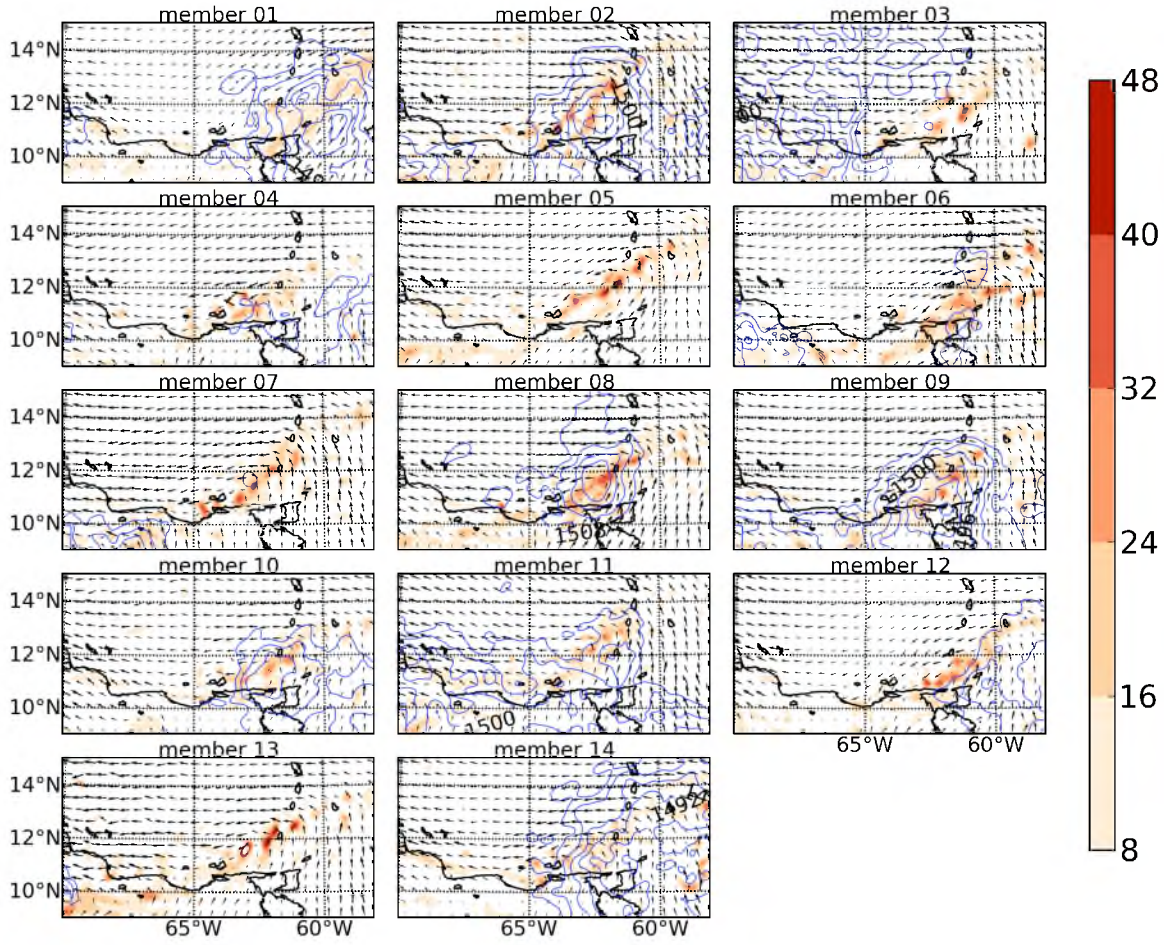

Figure 74. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of BS_CNTL ensemble members at 850 hPa at 1800 UTC 24th.

Table 13

Genesis statistics for

Hurricane Ernesto (forecast began 0000 UTC 24th)

Ensemble experiments	18Z 24th	12Z 25th
CNTL	07/14	13/14
BS_CNTL	04/14	N/A
BS_VERT	04/14	N/A
BS_PSIDOUBLE	09/14	N/A

3 and 14 in the BS_CNTL ensemble (Figure 74) compared with similar, or any, members in the CNTL ensemble (Figure 9). These unique geopotential height patterns may relate to the fact that SKEBS is adding energy in terms of θ, which closely relates to geopotential, and do not appear to coincide with unphysical features in other fields.

Comparing SKEBS ensembles with (BS_VERT; Figure 75) and without (BS_CNTL) a random phase vertical pattern generator, we find that at genesis time, there is not any systematic synoptic difference either in terms of variability or mean TC intensity. Both ensembles have members that predict intense TCs by this time as well as members whose disturbances are more similar to ordinary easterly waves; both ensembles have 4 of 14 members exhibiting TC genesis at this point (Table 13). In order to evaluate the impact of the streamfunction perturbation parameter in SKEBS, we doubled the amplitude of this psi parameter from 1.0E-5 to 2.0E-5 (Table 12) while leaving everything else equal (Figure 76). What we find is the fact that the enhanced amplitude in terms of streamfunction backscatter produces notably stronger tropical disturbances than does the BS_CNTL ensemble. This is especially true in terms of geopotential height anomalies, but also manifests itself in terms of tighter circulation patterns and elevated vorticity levels. Indeed, it makes sense that stronger TCs may result from higher amplitude energy backscatter perturbations in terms of the rotational wind component. As mentioned by Charron et al. (2010), perturbations to this particular wind component have particularly large inverse (i.e., upscale) cascade effects that can quickly manifest themselves in other aspects of the simulation. The ultimate result can be seen in the fact that, in this BS_PSIDOUBLE ensemble, 9 of 14 members show genesis, compared with just 4 of 14 members in the other two SKEBS ensembles (Table 12).

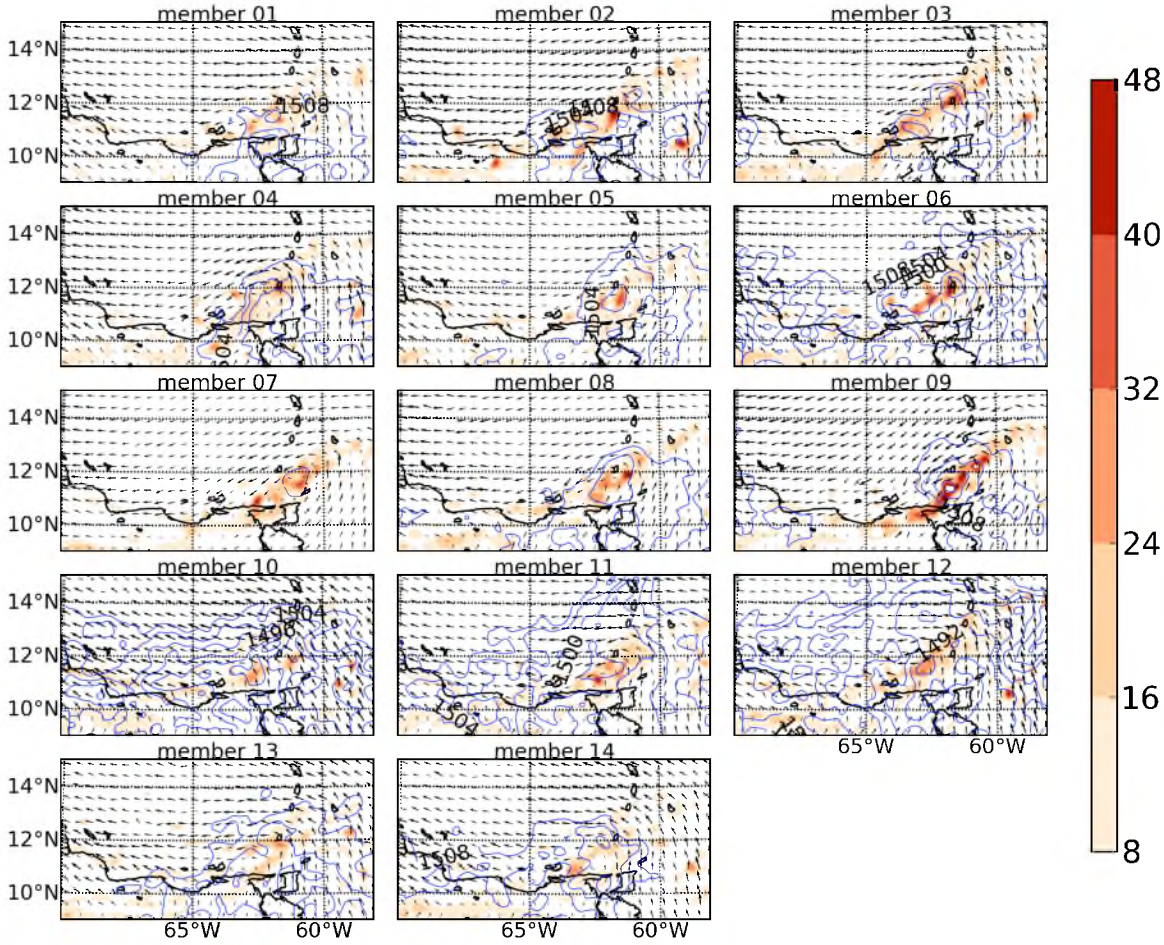

Figure 75. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of BS_VERT ensemble members at 850 hPa at 1800 UTC 24th.

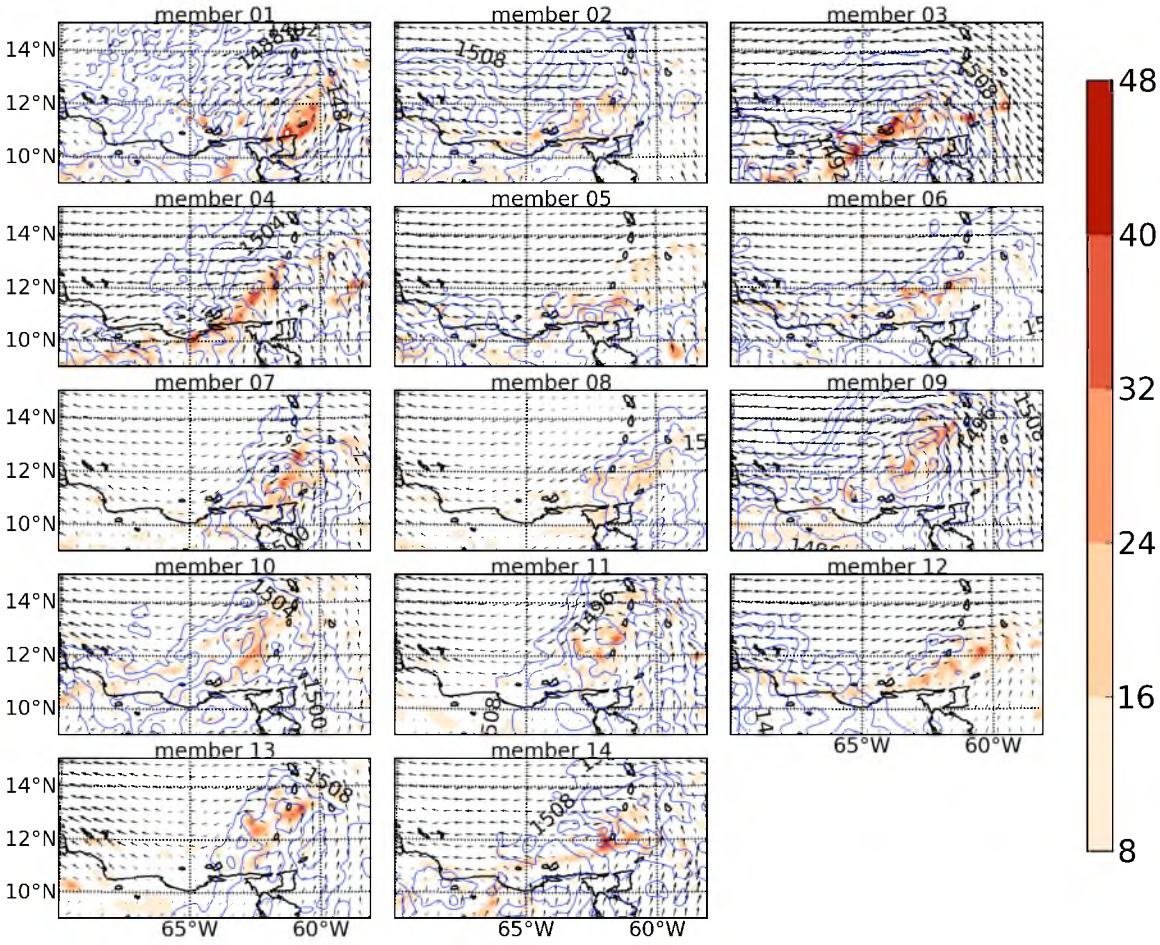

Figure 76. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of BS_PSIDOUBLE ensemble members at 850 hPa at 1800 UTC 24th.

Discussion

While it is difficult for ensemble studies related to the particulars of model error to draw clear, definitive conclusions, we feel that carefully constructed SKEBS ensembles provide many instructive comparisons against our control and breeding ensembles. In addition to an examination of these differences, we have also provided an analysis of the SKEBS scheme's performance characteristics as they relate to the SKEBS-related peculiarities and benefits found in the literature. For example, we confirmed that the heightened low-level θ bias found by Charron et al. (2010) does manifest itself quite clearly in the WRF-based SKEBS forecasts. This was peculiar, as the water vapor bias in the same regions was notably lower than that for the other ensembles we examined, and because of the fact that our SKEBS ensembles performed quite well across several other metrics.

In terms of the spectral energy differences, we found that the differences between error and spread in our SKEBS ensembles was notably lower than it was in our breeding and CNTL ensembles. This applied not only to the values found across the general domain, but also when considering intermediate and TC-proximate values. When considering spectral patterns in a 1000 km-per-side box centered on Ernesto, we found that spread values for all ensembles matched those of the error much more closely than they did when considering non-TC-proximate values. When approaching a TC, the spread/error gap actually closed most quickly for our SKEBS-based ensembles compared with the breeding and CNTL ensembles. Thus, it appears that the SKEBS scheme is better able to translate baroclinic and convective related error growth into ensemble spread for our particular case.

We find that, in general, values for E-dimension spread variance are occasionally just as high as they are for error variance. This is true particularly for breeding ensembles over much of the simulation period. However, when considering values within 200 km of the TC center, the E-dimension for spread variance is consistently lower than that for error variance. Thus, we find that, near TCs, the ensembles have a more difficult time translating error variance dimensionality into spread dimensionality, which is opposite of our related finding in terms of spectral energy.

In terms of the ability of the SKEBS control ensemble to produce genesis, we find that it has a more difficult time than does the ET-based control ensemble. While the vertical random pattern generator option in SKEBS does not produce any more genesis cases than the SKEBS control, doubling the streamfunction perturbation forcing does result in more TCs. These results makes sense, as this parameter change effectively adds more energy at the small parts of the spectrum in terms of the rotational wind, which component not only provides heightened upscale energy cascades (Charron et al. 2010), but also affects small-scale vorticity directly.

Highlights of Chapter 8

- SKEBS ensembles show notably higher spread, similar error, and thus less underdispersion compared to CNTL and breeding ensembles.
- Near the simulated TCs, underdispersion is lower for all ensembles, but it is particularly improved for the SKBES ensembles. It thus appears that the SKEBS scheme is better able to translate baroclinic and convective related error growth into ensemble spread for our particular case.
- In terms of the ability of SKEBS to reduce spectral slope, we find that

compared to CNTL, this occurs *above* 80 km wavelength because of heightened spread maximized *near* 80 km wavelength.

- SKEBS forecasts show heightened low-level θ bias, as noted by Charron et al. (2010). By contrast, the same backscatter forecasts showed notably low vapor bias.
- In terms of genesis, we find that the BS_CNTL ensemble forecasts similar strength TCs as CNTL does at genesis time, but that the SKEBS scheme produces notably higher spread in terms of intensity and spatial positioning, which is a beneficial result, and likely relates to beneficial spectral perturbation changes as well.

CHAPTER 9

AN ADDITIONAL CASE STUDY WITH TYPHOON NURI

In order to better determine, as far as is possible, the robustness of our results, several of our previous perturbation methods are used to create three ensembles that forecast the genesis of Typhoon Nuri (2008), which was known for being notoriously difficult to forecast (Snyder et al. 2011). Nuri originated from a "finite-amplitude wave/vortex structure" which tracked westward in the west Pacific several days before the disturbance reached tropical depression (TD) status around 1800 UTC 16 Aug 2008 (JTWC; see also Montgomery et al. 2010). Two days before genesis, NOGAPS ensembles, created as part of the T-PARC/TCS-08 field campaigns, showed a highly uncertain forecast, as various of the 32 ensemble members predicted anywhere from full-fledged genesis to nondevelopment (Snyder et al. 2011)

Similar to the Ernesto case, our forecasts of Typhoon Nuri will begin 1 day before genesis and will be examined using the FNLq3, FNLuvqt3, and BS_CNTL ensemble configurations (Table 2 and 12). Also, similar to the previous chapters, the breeding simulations are performed using WRFV3.3.0 while the backscatter ensemble is implemented in WRFV3.4.1. The breeding scheme here runs from 0000 UTC 15th to 0000 UTC 16th Aug 2008. Subsequently, the simulations are run from 0000 UTC to 1800 UTC 16th Aug 2008, which latter time is when TC Nuri becomes a tropical depression according to the JTWC. The domain setup is detailed in Figure 77 and Table 14.

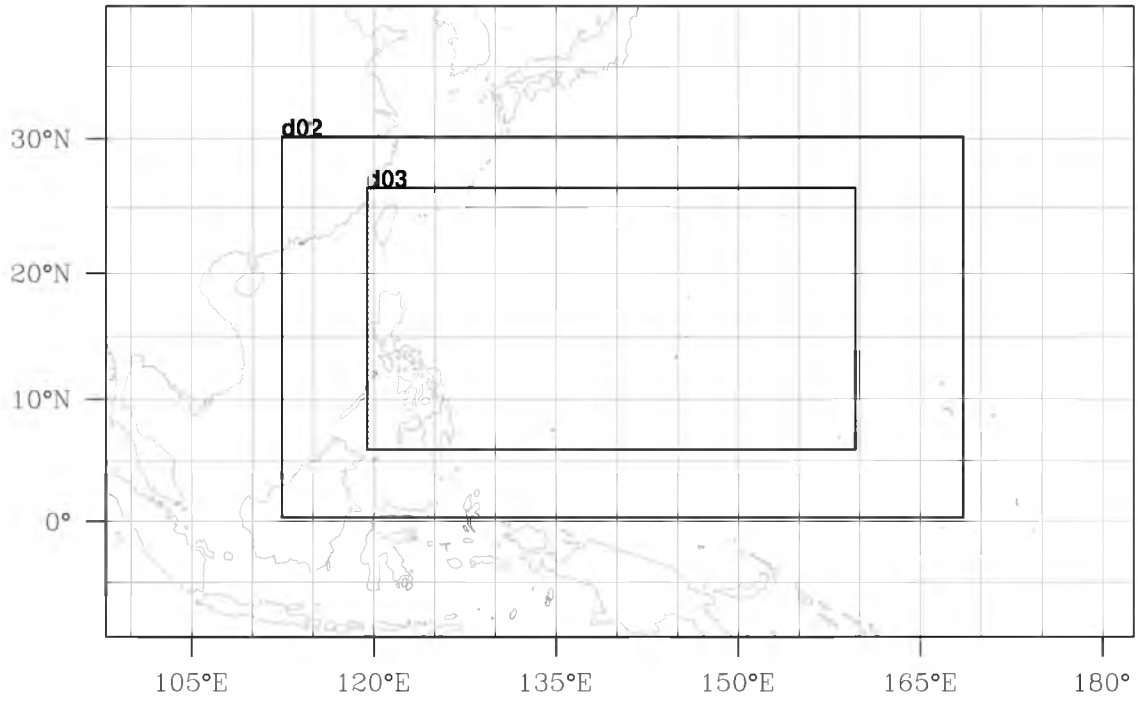

Figure 77. WRF domains used for all ensemble simulations of Hurricane Nuri.

Table 14

Dimensions, grid spaces, and time steps for model domains in Nuri simulations

Domain	Dimension (x × y × z)	Grid space	Time step
1	182 x 112 × 36	45 km	120 s
2	361 x 202 × 36	15 km	40 s
3	772 x 412 × 36	5 km	13.3 s

Basic evaluation

The analysis of the Nuri results loosely follows those of TC Ernesto, wherein we contrast the various ensembles in terms of bias, IC perturbations, the synoptic features, and environmental precursors relating to genesis. To begin, we will analyze the bias of these ensembles over the entirety of the free simulation period. While we examine post-breeding initial condition differences later, and thus here are not moving in strict chronological order, we provide the bias over time first as it renders a quick check of the plausibility of the ensembles' results. Starting near the surface (Figure 78), we find that there are significant bias differences between the Nuri ensembles and those of Ernesto. These arise quickly, as we find the FNLq3_NURI ensemble starting with notably higher vapor bias at 925 hPa than any previous ensemble in this study. While we find that, over time, the backscatter ensemble does provide the lowest moisture bias of any of the Nuri ensembles, all three of these Nuri ensembles generally show notably higher bias than the ensembles forecasting Hurricane Ernesto. In terms of θ (Figure 79b), we find that the pattern is similar, in that the error is generally higher for the ensemble forecasts of Nuri, while the BS_CNTL_NURI ensemble performs the best of the three at 925 hPa.

Moving to 850 hPa (Figure 79b), we find that the Nuri results confirm what was found by Charron et al. (2010), in that the SKEBS low-level θ bias is often higher than other ensemble techniques. Note how much the SKEBS θ bias degrades from 925 to 850 hPa. In terms of *vapor* at 850 hPa, we find that the backscatter technique provides notably *lower* bias than the FNLq3_NURI ensemble. At 700 hPa (Figure 80), the poor performance of SKEBS in terms of θ disappears and the ensemble again compares favorably to the others, although not quite as favorably as the BS_CNTL_NURI vapor

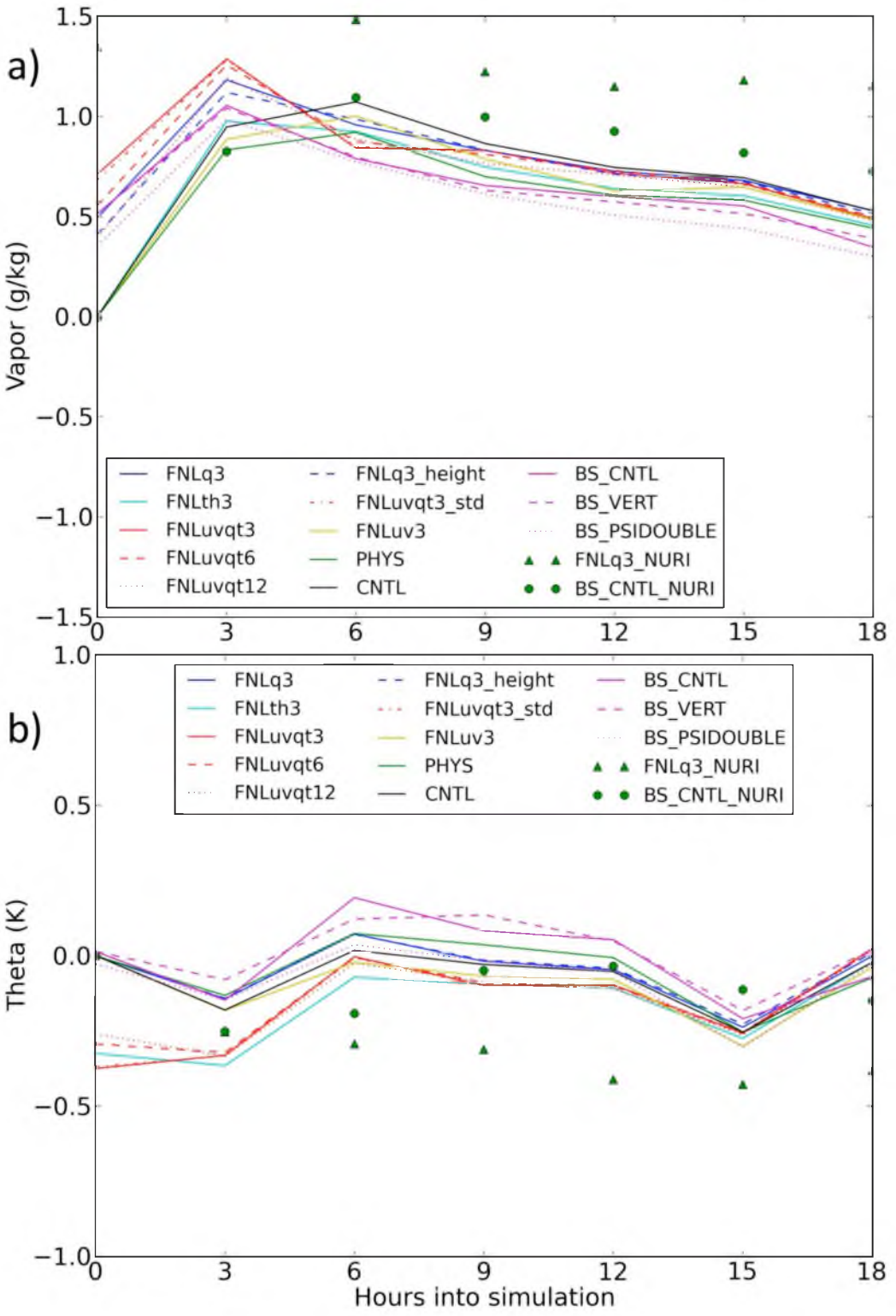

Figure 78. Comparison time series of bias at 925 hPa between Nuri and Ernesto ensembles for a) vapor (g/kg) and b) theta (K).

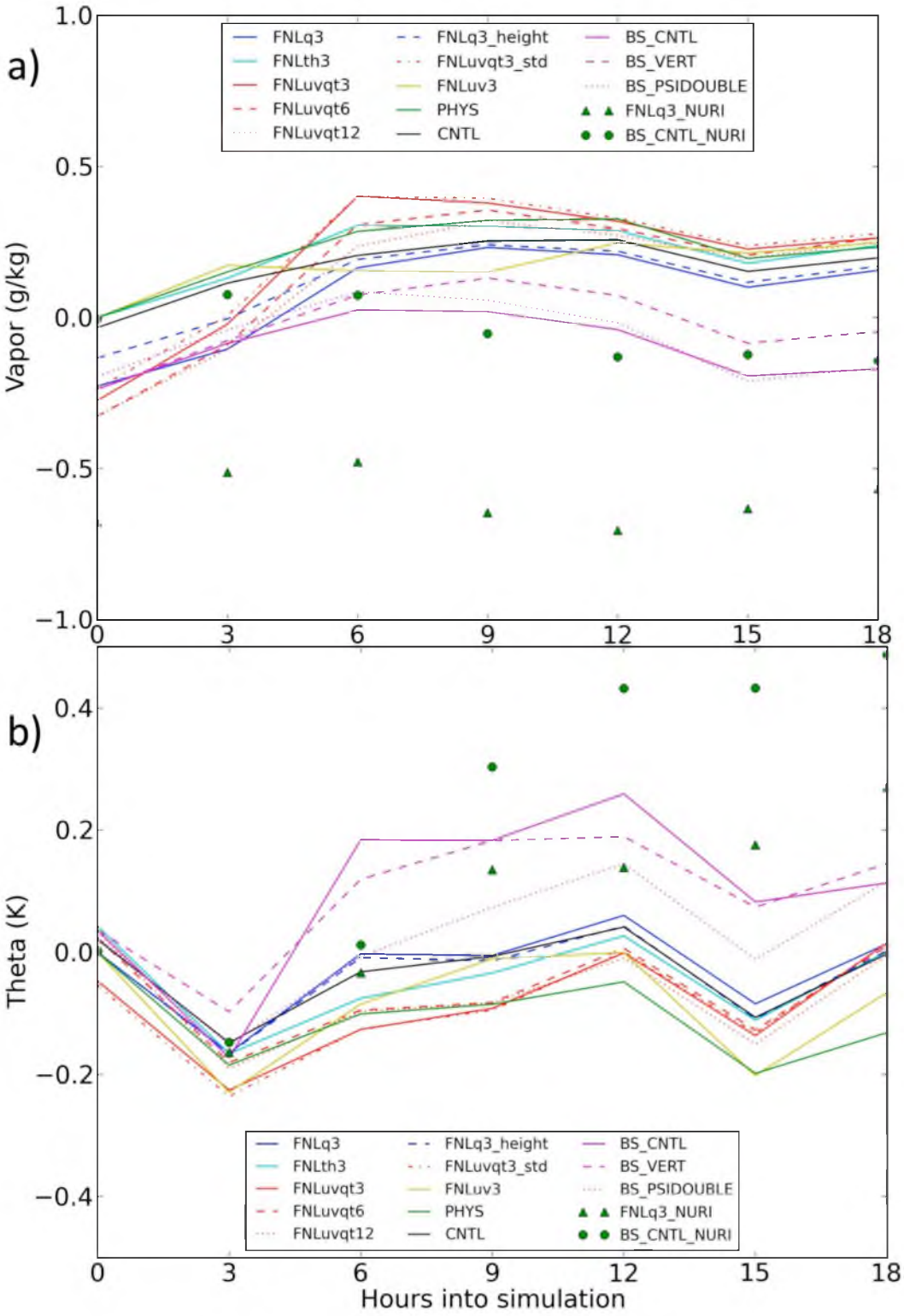

Figure 79. Comparison time series of bias at 850 hPa between Nuri and Ernesto ensembles for a) vapor (g/kg) and b) theta (K).

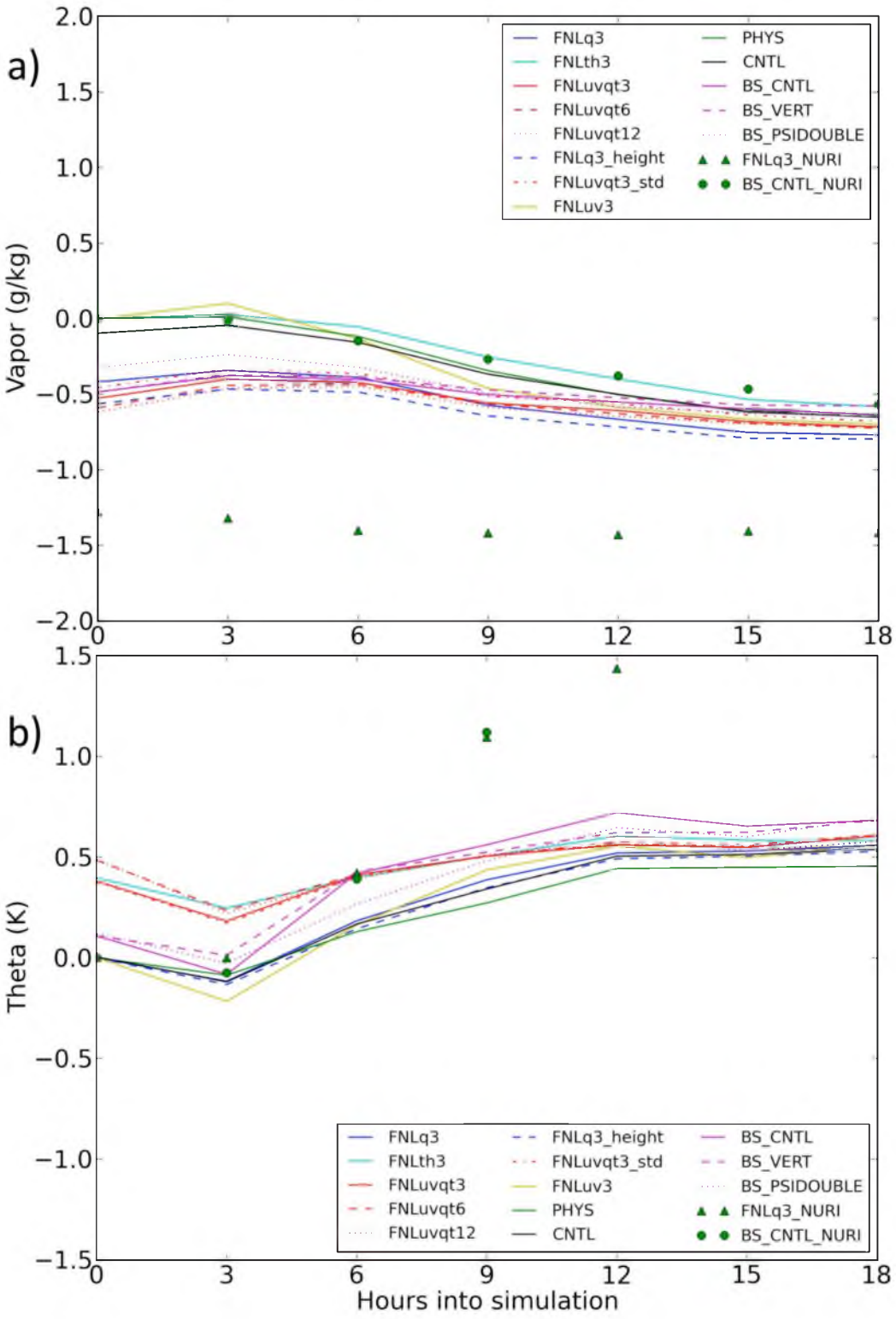

Figure 80. Comparison time series of bias at 700 hPa between Nuri and Ernesto ensembles for a) vapor (g/kg) and b) theta (K).

field. Comparing backscatter θ bias behavior with that of the other ensembles, for both Ernesto and Nuri, we find that the notably poor performance is confined to 850 hPa (Figure 79). At 500 hPa (Figure 81), we find that the Nuri θ bias is generally comparable to that of Ernesto, while the Nuri vapor bias is notably higher thanthat of Ernesto.

Figure 82 shows these values averaged from 1200 to 1800 UTC at all model heights. It not only confirms our previous observations, but provides a simplified way of noting that well into the simulation period, the SKEBS θ bias for Nuri does perform worse than the other Nuri ensembles. At this time, this not only occurs around 850 hPa, but actually from there and above. In the magenta lines of Figure 82b, one sees that the corresponding data for Ernesto are quite similar, but that this SKEBS θ underperformance occurs only up until around 350 hPa. This poor performance is all the more surprising because it is so different from the vapor comparisons. Note that the BS_CNTL_NURI vapor bias is often drastically reduced compared to that of the FNLq3_NURI and this occurs for the SKEBS ensembles of Ernesto as well. Overall, the general bias of the Nuri ensembles is notably higher than that of the Ernesto ensembles, which is what we would expect to occur if Nuri were a more uncertain forecast. Also important here is the fact that our ensemble creation methods produce forecasts of Typhoon Nuri with reasonable vapor and θ bias in a difficult predictive environment. In order to view the differences in final perturbation patterns produced between breeding forecasts of Nuri and Ernesto, we subtract the analysis from the predicted values and plot the resulting 850 hPa water vapor bias for each ensemble member for FNLq3_NURI at 0000 UTC 16th Aug 2008 (Figure 83). Notice that while the large water vapor deficit across the center part of the

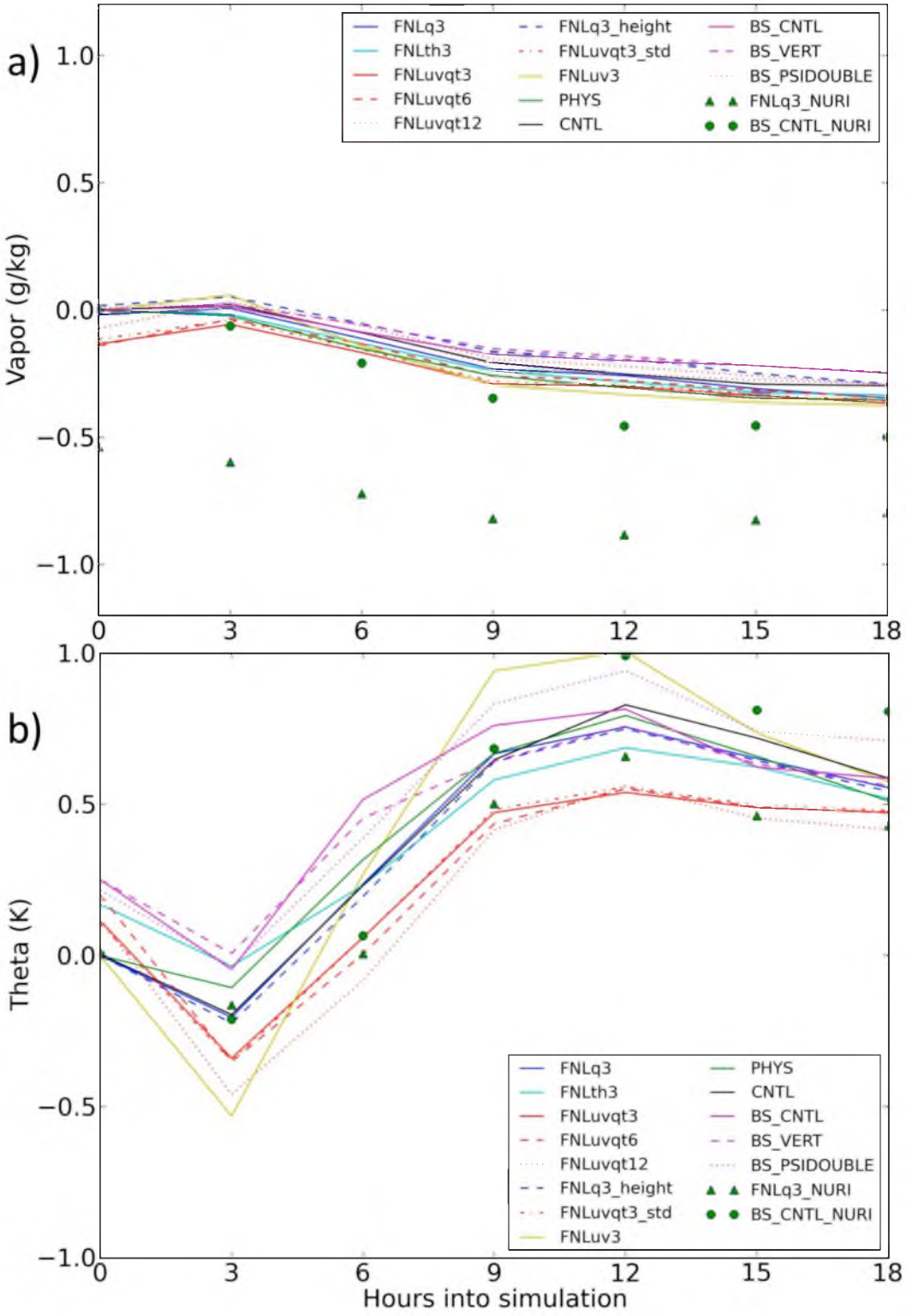

Figure 81. Comparison time series of bias at 500 hPa between Nuri and Ernesto ensembles for a) vapor (g/kg) and b) theta (K).

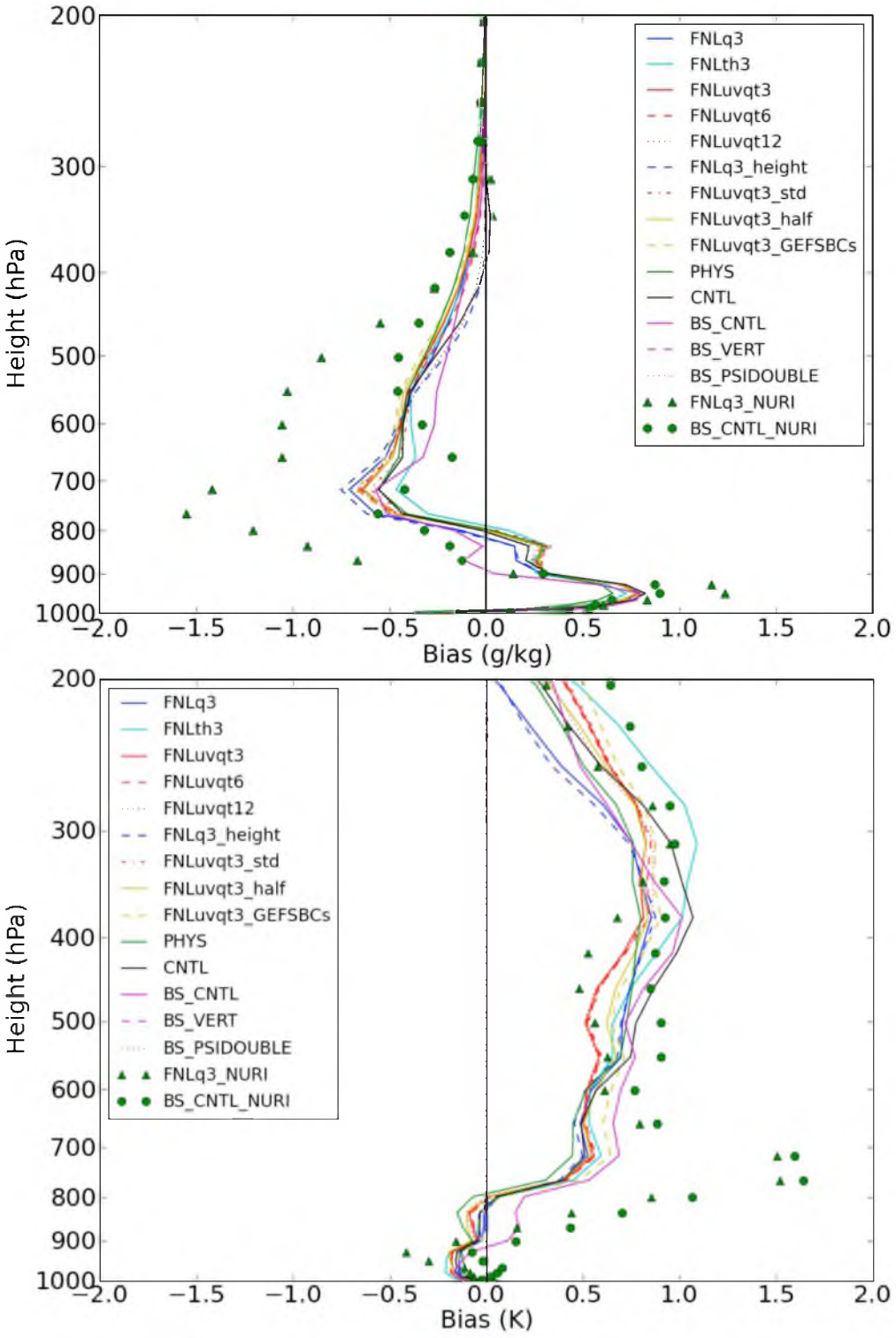

Figure 82. Comparison of bias by height from 1200-1800 UTC 24th August between Nuri and Ernesto ensembles for a) vapor (g/kg) and b) theta (K).

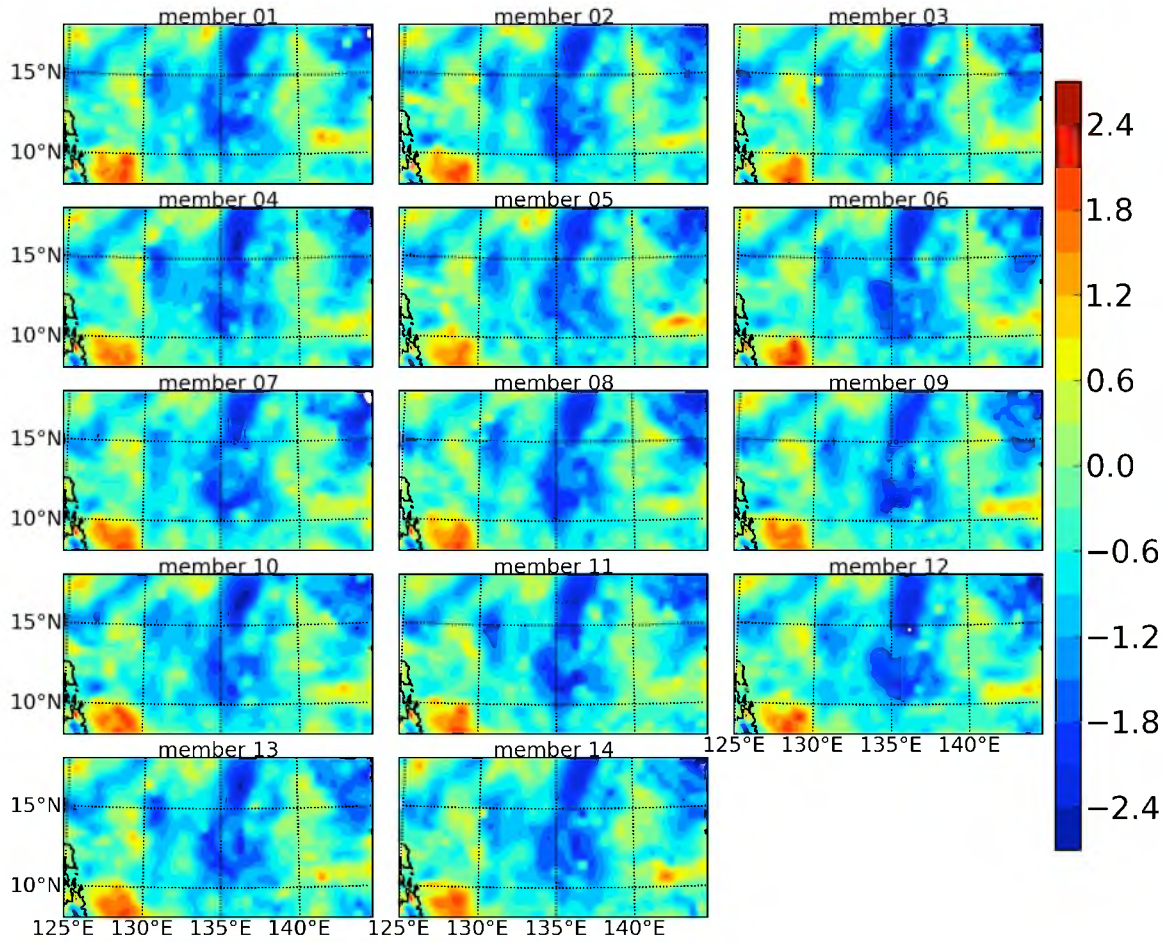

Figure 83. Bias of 850 hPa vapor (g/kg) for each member of FNLq3_NURI ensemble at 0000 UTC 16th (end of breeding period).

domain is present in all members, there is a notable amount of variability in the small-scale structure. This is similar to the corresponding Ernesto forecast (FNLq3; Figure 15), where we note a general lack of moisture in the southeast corner with the accompanying small-scale variability. Also note that the scales are the same in Figure 15 and 84, all of which helps demonstrate that the breeding scheme as used in the FNLq3 ensembles is robust to the extent that it provides reasonable initial perturbations across more than just one TC case.

Plotted in Figure 84 is the same initial plan view of vapor bias, but for FNLuvqt3_NURI. Note that the *intermember* differences are similar to the FNLq3_NURI ensemble (Figure 83), but that there are notable differences in the large-scale patterns, focusing on the same member, between the two ensembles. For example, the FNLuvqt3_NURI members notably reduce the dry bias in the northern part of the domain around 135°E compared with the FNLq3 members. This is similar to the differences between the FNLq3 and FNLuvqt3 ensembles (Figures 15 and 27), wherein the FNLuvqt3 notably reduces the significant dry bias found in the southeast corner of FNLq3. Overall, however, in the bred forecasts of Nuri, the initial vapor perturbation patterns are of similar magnitude as the corresponding forecasts of Ernesto, yet the particular pattern differences between the members are quite unique to the underlying flow of the particular TC of interest. In addition, the bred ensembles of Nuri appear to show notable dry biases in critical layers, which, based on what we have seen occur in forecasts of TC Ernesto, may significantly affect the likelihood of forecasts of TC genesis.

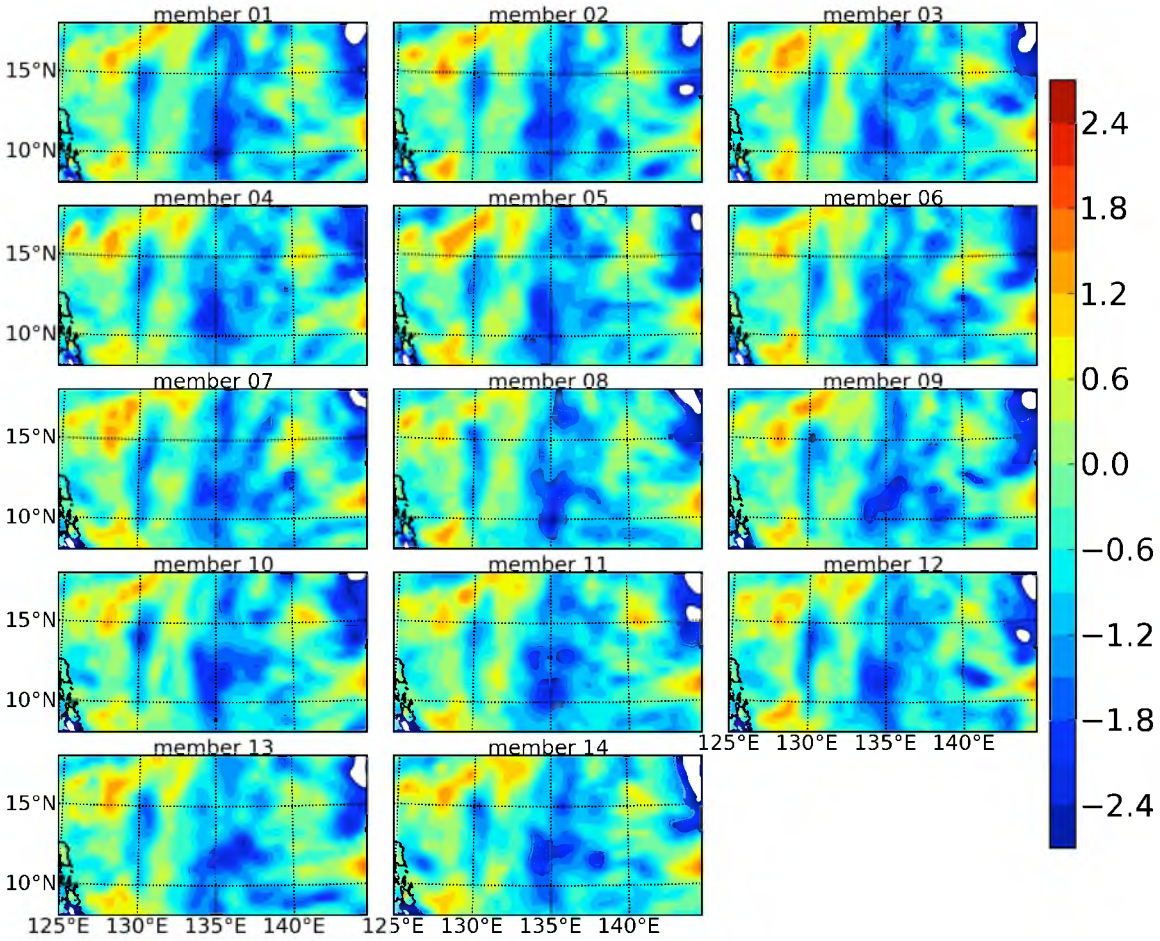

Figure 84. Bias of 850 hPa vapor (g/kg) for each member of FNLuvqt3_NURI ensemble at 0000 UTC 16th (end of breeding period).

Genesis results

To understand the basis from which these ensembles are created, we plot in Figure 85 a synoptic TC view from the Nuri deterministic forecast based on FNL ICs and BCs, at the time of actual genesis (1800 UTC 16th Aug 2008). What we find is that while this forecast shows a fairly symmetric cyclonic circulation, the disturbance is quite weak compared with the deterministic forecast of TC Ernesto (Figure 51). Also notice the fact that, compared to Ernesto, Nuri shows notably lower vorticity, which is oddly accompanied by *reduced* geopotential height, which reminds us of the fact that physical and environmental factors related to TC formation differ slightly from the Atlantic (Ernesto) to the Pacific (Nuri). For example, in the Pacific, genesis is more related to tropical upper-tropospheric troughs, mesoscale convective systems (Harr et al. 1996), and the Madden-Julian oscillation (Slade and Maloney 2012), whereas in the Atlantic, storms typically form as a result of African Easterly Waves. In addition, the Pacific is not as well observed as the Atlantic. At this stage, however, we will only concern ourselves with the fact that the deterministic forecast of TC Nuri produces a notably weaker disturbance, one day ahead, than does a similar forecast of TC Ernesto. From this basis, we will be able to better analyze the effect of the various perturbation methods on the ensemble results.

Using a similar synoptic style as found in our earlier discussion, we plot 850 hPa vorticity and geopotential height for all 14 members of the FNLq3_NURI ensemble in Figure 86.What we find is that the vapor breeding and finalized vapor perturbations do not produce notable differences among the ensemble members, which causes most of the ensemble's disturbances to be weak, similar to that of the deterministic forecast (Figure

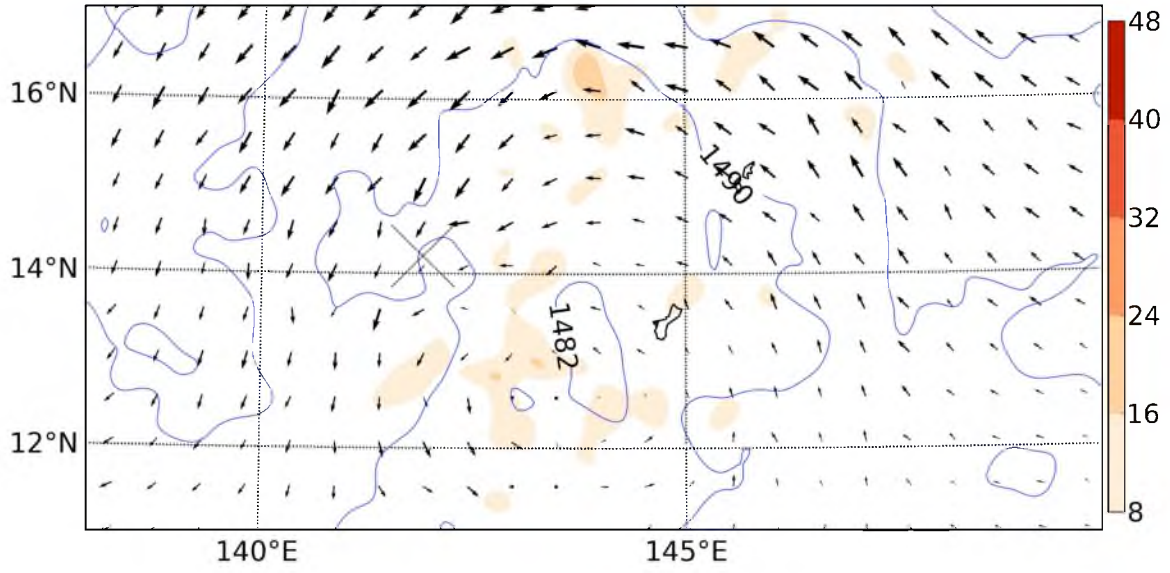

Figure 85. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of Hurricane Nuri at 850 hPa at 1800 UTC 16th from WRF deterministic forecast using FNL ICs and BCs.

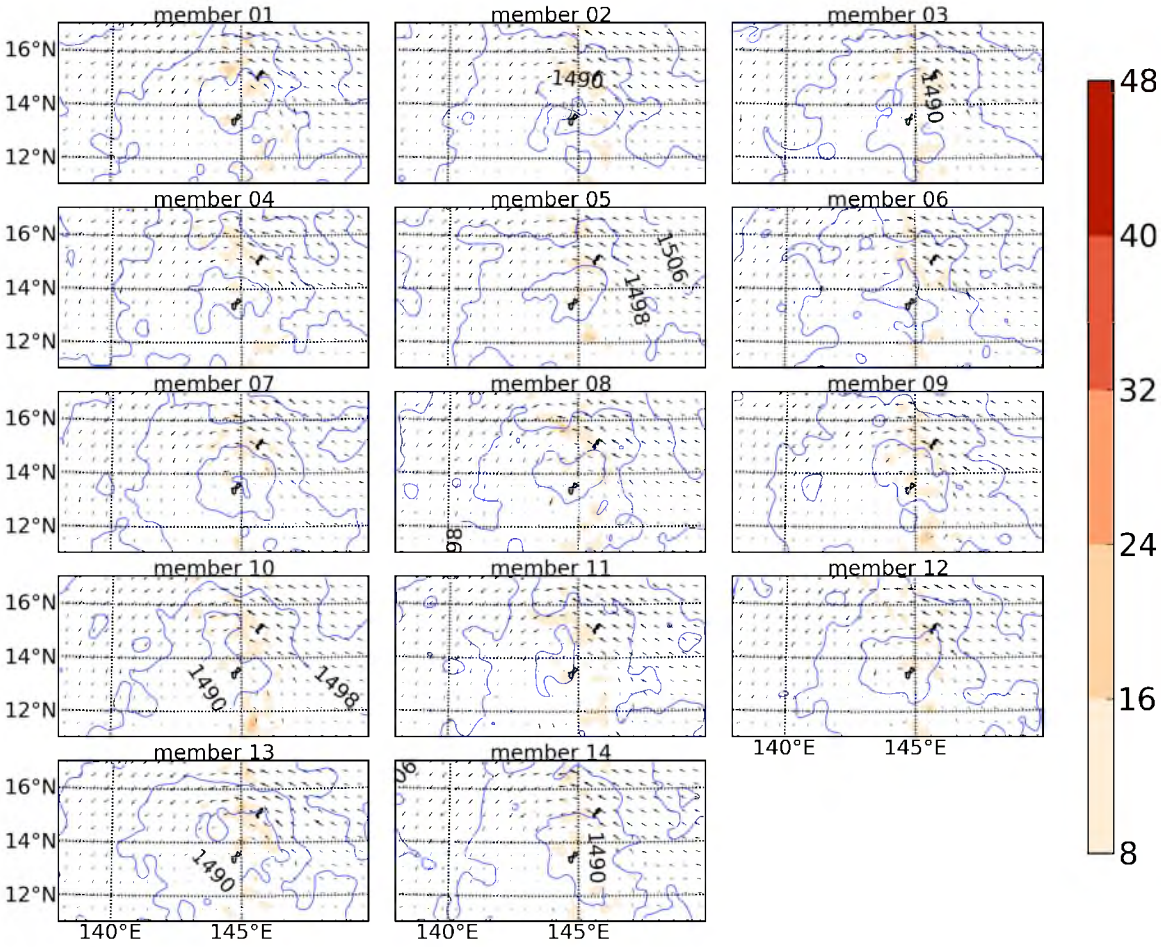

Figure 86. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLq3_NURI ensemble members at 850 hPa at 1800 UTC 16th.

85). Despite the different circumstances between Nuri and Ernesto, we find that each ensemble behaves fairly similarly to its corresponding deterministic simulation. Note, for example, how Ernesto's FNLq3 ensemble (Figure 10) also shows similar deviations, in terms of intensity or spatial positioning, from its deterministic simulation (Figure 51).

Our objective in performing the simulations of TC Nuri was to test the robustness of our previously-discussed ensemble methods. Seeing that breeding one variable versus four produced notably disparate results (notice the differences between Figure 10 and 30, for example), we also chose the FNLuvqt3 ensemble technique to test with our Nuri case. Recall that, following McLay et al. (2008), we use a spatially invariant scaling factor in our breeding simulations. Essentially, this assumes similar error growth across space, and is justified here and in their paper by the fact that we stay within the tropical error growth regime for our *mesoscale* ensemble simulations. This was in response to the fact that midlatitude regions often exhibit notably different error growth characteristics compared with those of the Tropics (Zhang and Krishnamurti 1997). While our justification worked well in our forecasts of the Atlantic-based Ernesto, spatial variability in error growth rates became an issue in the FNLuvqt3_NURI ensemble. This is largely due to the particular circumstances surrounding θ breeding. While in Ernesto we found that θ had notably higher error growth rates than did vapor, or either component of the wind, here we discovered that near the top of the model in the FNLuvqt3_NURI simulations, there was a notable difference in θ error growth rates with height. This difference was so large that the scaling factor used in the process was almost entirely influenced by the inordinate amount of θ error growth occurring in the highest levels of the atmosphere. Essentially, the lower level perturbations were being rescaled according to the rapid error growth

much higher up, and consequently, the low-level values of θ in the various members converge during the breeding period. The result of this is seen in Figure 87, where the members are all very similar in terms of vorticity and geopotential height, and none produce TC genesis due to the problem with the invariant θ scaling factor; in fact, despite being unrealistic, the forecast results in the figure are included to demonstrate the potential effects of this spatially-invariant scaling factor.

We did attempt to correct for this difference in θ error growth rates with height, and created a corresponding vertical localization ensemble using the technique discussed earlier in this dissertation. Again, this entails using four error scaling factors, split by height, such that perturbations for each member are maintained to a reasonable extent for that particular level. While this localization better maintains perturbation size in the lower levels compared with FNLuvqt3_NURI, there appear to be too few vertical levels used for rescaling, and eventually, θ differences among the members in the lower levels converge. Computational resources at the moment inhibit tests with significantly more height-specific scaling factors. However, future work involving a more discrete height localization will help us determine the effectiveness of the breeding scheme in the face of sharp spatial error growth differences through the simulation.

Lastly, we run a SKEBS-based ensemble of TC Nuri in order to examine the effect of energy backscatter on the resultant genesis likelihood and intensity spread. Similar to BS_CNTL which was performed for Hurricane Nuri (Table 12), here we use the standard amplitudes of θ and streamfunction perturbation and turn off the SKEBS vertical structure random pattern generator. Vorticity and geopotential heights at 850 hPa for each of the corresponding members are plotted in Figure 88. What we see is that, compared

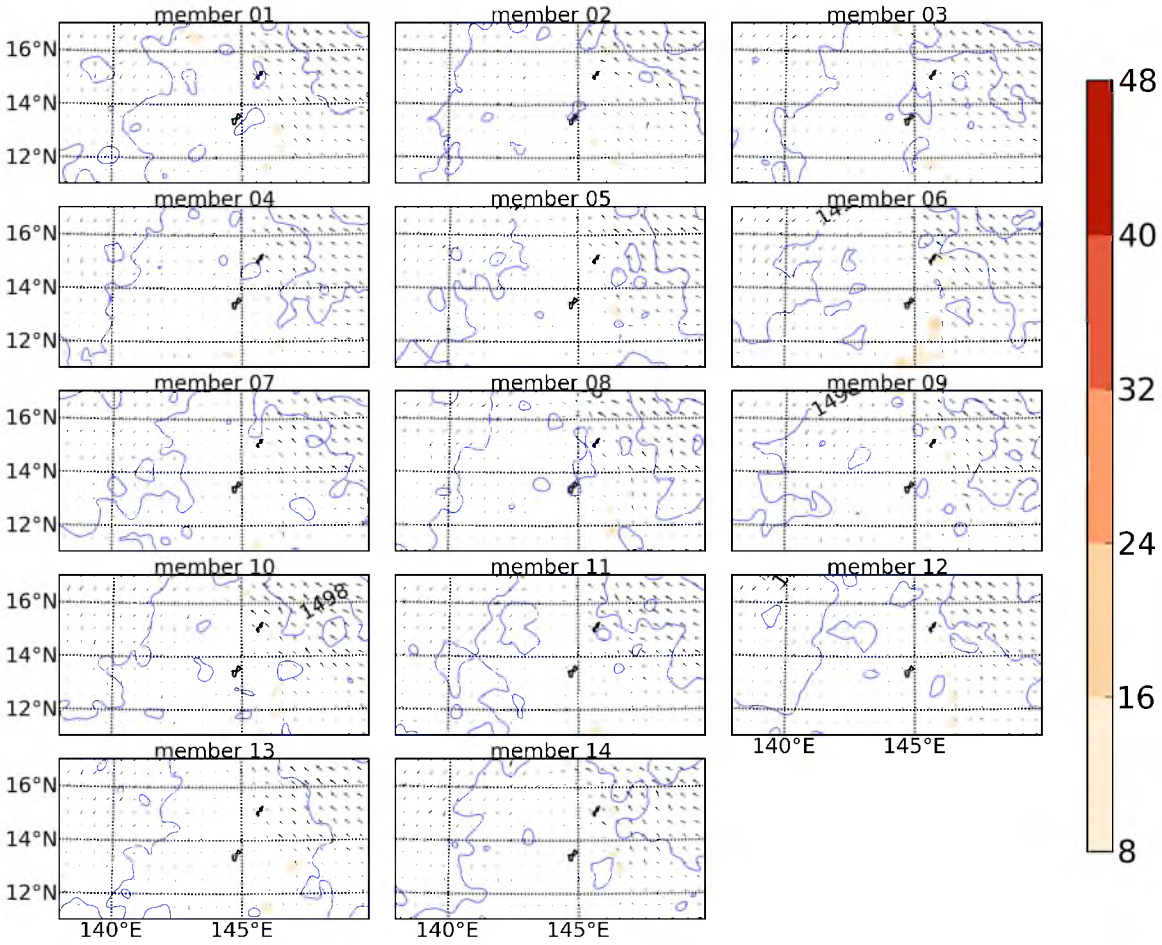

Figure 87. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of FNLuvqt3_NURI ensemble members at 850 hPa at 1800 UTC 16th.

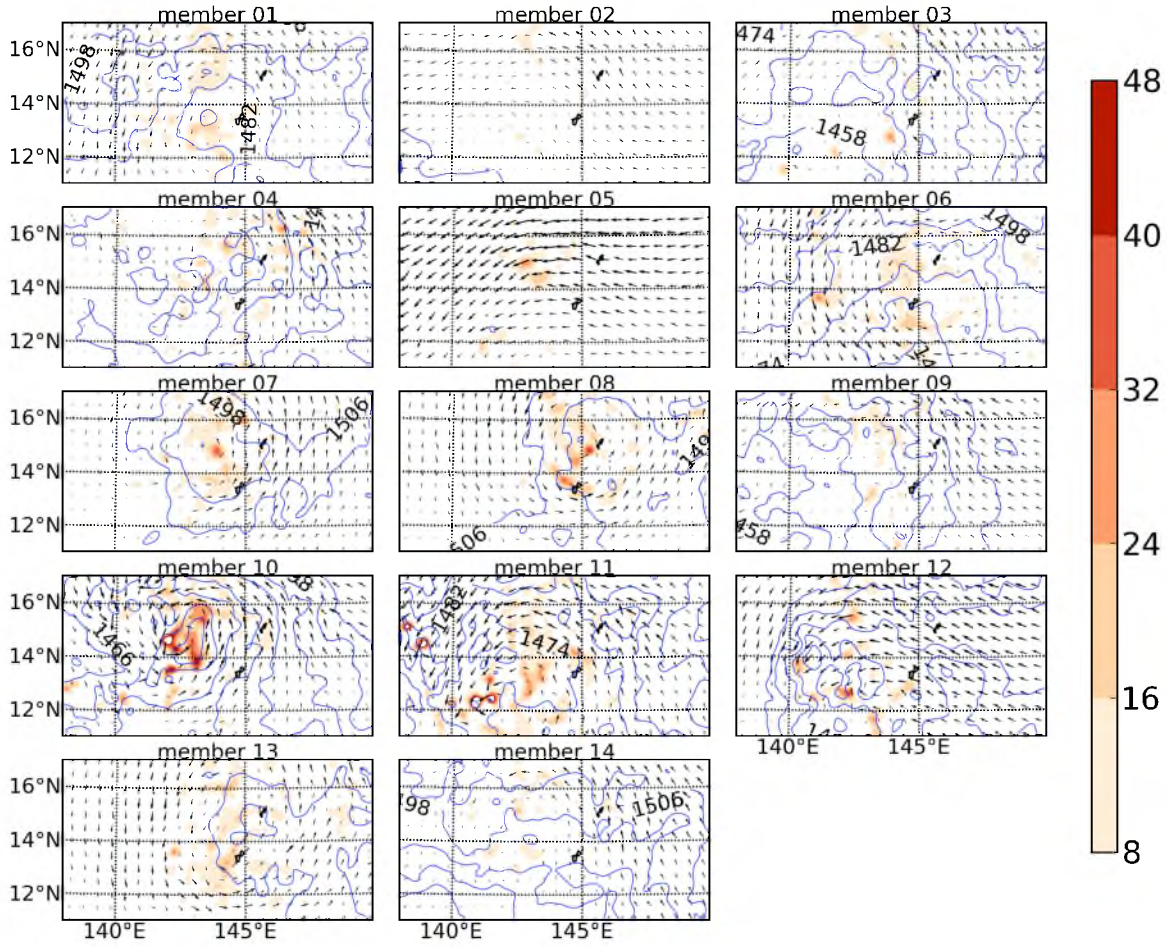

Figure 88. Vorticity (shaded; * 10^{-5} s^{-1}) and geopotential height (contours; 4 m intervals) of BS_CNTL_NURI ensemble members at 850 hPa at 1800 UTC 24th.

with our breeding ensembles, which appear to suffer from low levels of water vapor and spatially disparate error growth, the BS_CNTL_NURI ensemble shows several intense storms while, at the same time, providing a large amount of intensity variance among the members. While members 8 and 10 show well-organized, fairly intense TCs, other members (2 and 5) show almost no closed cyclonic circulation at all, and other members fall at various points in between. Note also the large amount of spatial variability, manifest in the fact that many of the members produce TCs hundreds of miles from each other. Such uncertainty is an essential characteristic of ensembles forecasting highly-uncertain TC genesis cases.

Overall, in these simulations of Typhoon Nuri, we find that the breeding and SKEBS schemes largely produce realistic results despite the large differences from Hurricane Ernesto. While rapid upper-level θ error growth inhibits a full evaluation of FNLuvqt3_NURI, we find that the FNLq3_NURI and BS_CNTL_NURI ensembles behave quite similarly to those of Hurricane Ernesto. This is not only in terms of bias, but also in terms of spread between the members, the ability to forecast genesis, and perturbation spatial characteristics. We also learn that in the breeding scheme, the spatially invariant scaling factor must be used carefully even when confined to the Tropics. Instead of finding latitudinal error growth concerns, as addressed by Mackey and Krishnamurti (2001), when breeding four variables, we instead discovered sharp error growth changes with height, which resulted in notably small spread near the surface. Finally, the SKEBS scheme looks particularly promising, considering the number of strong TCs it forecast in our ensemble of both Ernesto and Nuri. At the same time, the SKEBS technique provided plausible estimates of uncertainty, as manifest in the notable

TC intensity spread among its ensemble members.

Highlights of Chapter 9

- Vapor breeding forecasts of Typhoon Nuri show notable similarities with corresponding forecasts of Hurricane Ernesto, in terms of spatial perturbation patterns and spread.
- Results from the FNLq_NURI ensemble confirm that, while water vapor is important to TC genesis, perturbations of the field, on their own, are not sufficient to create a realistic ensemble.
- The SKEBS forecast of Nuri shows a remarkable amount of spread among the members in terms of TC intensity, along with several forecasts of well-organized TCs. This method is unparalleled in our study in terms of its ability to forecast both intense TCs and completely dissipative cases in the same ensemble.
- The simulations of Typhoon Nuri provide more evidence that Charron et al. (2010) is correct that the SKEBS scheme produces notably high θ bias in the low-levels (specifically, around 850 hPa), while producing remarkably *low* bias in terms of water vapor.
- While reasonable, bias for both water vapor and θ are notably higher for forecasts of Nuri compared with Ernesto, for all ensembles used. While this difference is expected, its quantification does provide a more realistic idea of the benefits and drawbacks associated with regional ensemble forecasts of TC genesis.

CHAPTER 10

SUMMARY AND CONCLUDING REMARKS

Using various ensemble configurations, we have examined TC genesis forecasts, error growth characteristics, dimensionality, and spatial structures associated with high-resolution forecasts of Hurricane Ernesto. We found that a high-resolution regional ensemble (CNTL), downscaled from the Global Ensemble Forecast System (GEFS), forecasts TC genesis with a higher likelihood, 1 day ahead, compared with the GEFS forecasts made by NCEP. By best track genesis time, the mesoscale forecast predicts 7 out of 14 members as showing genesis, compared with 4 of 14 in the GEFS simulations. Thus, using a higher resolution ensemble appears to help overcome the difficult forecast problem Hurricane Ernesto posed before its genesis. Depending on the configuration, breeding ensembles may produce notably stronger TCs than the CNTL forecasts of Hurricane Ernesto. However, the dispersion between these BGM members is often low compared to that of CNTL. Using these BGM ensemble forecasts, the effects of perturbing one or several principal variables on key ensemble characteristics have been quantified.

In terms of the differences between the breeding ensemble error growth rates, that of θ was consistently higher than that of both water vapor and u and v wind components. No matter which of these variables we bred, error growth saturates 12 to 18 h into the 24 h breeding period. This 12 to 18 h error growth rate plateau helps confirm that our choice

to breed for only 24 h is reasonably in line with the underlying forecast error growth. In terms of the perturbation differences between ensemble members at the end of the 24 h breeding period, in general, it was found that the θ field exhibited much more variability than did water vapor, which likely relates to the higher θ versus vapor error growth rates mentioned above. This higher intermember θ variability is not only true when comparing θ and vapor fields in FNLth3 and FNLq3, but also for those which are found in CNTL, which helps confirm the results of the simple breeding method here implemented.

Because of the associated error growth characteristics, at the end of the 24 h breeding period, the ensembles breeding water vapor ended up with up to 0.5 g/kg less domain-averaged moisture at 700 and 850 hPa compared with the ensembles which relied on the FNL analysis for their water vapor fields. It appears that because of these differences, the ensembles that did not breed vapor, whether by itself or with other variables, produced notably stronger TCs by 1800 UTC 24th Aug than those that did.

In examining the differences between 3, 6, and 12 h cycling periods, we found that the longer periods did indeed experience more error growth during the breeding period, but that this growth was not commensurate with the extended length of the cycle period. For example, because error growth quickly becomes nonlinear, changing from a 3 to 6 h cycle period only produced errors whose magnitudes were 0 to 50% higher. Thus, the 3 h breeding cycle produces the best tradeoff in terms of capturing the most error growth per cycle while producing a large number of cycles. A larger number of cycles provides the scheme more chances to sample the different directions of underlying forecast error growth. The FNLuvqt12 ensemble ended up producing less intense TCs compared with the FNLuvqt3 and FNLuvqt6 ensembles, which might be attributed to the larger initial perturbations that may temporarily inhibit TC spin-up because of the field balances which first must occur.

We find that error dimensionality is much more volatile near TCs compared with values at large radii. This is manifest in a notable midsimulation drop in E-dimension within 200 km of the simulated TC centers and appears to relate to the fact that convective and baroclinic energy conversion reduces the local error dimensionality (Oczkowski 2005). Near the TCs, the ensembles bred with four variables did not experience the midsimulation low-dimensionality, although the reasons behind this are not clear. Perhaps relatedly, these FNLuvqt ensembles (especially FNLuvqt3) also had a relatively low correlation between their water vapor dimensionality and spread, compared to what was found in the FNLq3 and FNLth3 ensembles.

The vertical localization, when used, did produce smaller perturbations near the surface, and larger perturbations in the midlevels. Likely because of the reduced amount of water vapor from 500-800 hPa in the FNLq3_height ensemble near the initial disturbance, this ensemble produced weaker TCs than did the FNLq3 ensemble, which used a *vertically-invariant* scaling factor. Also, the vertical localization did not notably improve bias at 925, 850, 700 hPa compared with the nonlocalization ensembles. This may imply that, for high-resolution breeding forecasts of Hurricane Ernesto, a *vertically-invariant* scaling factor produces a favorable computational cost to accuracy tradeoff. The one notable benefit that we did find was that the localization caused the ensemble to have a notably higher correlation between its domain-averaged water vapor RMSE and spread.

We tested GEFS against FNL boundary conditions and found that the former produced notably higher θ bias above roughly 700 hPa. The vapor bias also was worse with the GEFS BC-based ensemble, but only slightly and from 400 to 700 hPa. The GEFS BC-based ensemble did, however, produce slightly stronger TCs without lowering the intensity variance among the members. When testing the importance of perturbation

size, we found that the FNLuvqt3_half ensemble produced notably higher θ bias compared with FNLuvqt3 above 600 hPa. While FNLuvqt3_half did produce strong TCs, there was literally no intensity spread among the members because they perturbations ended up very close to the deterministic simulation.

A multiphysics ensemble is created and we find that it produces strong TCs, but relatively little variance among the members. In addition, it shows notably higher error compared to CNTL and the breeding ensembles for the first 9 h of the simulation period. Based on the combined 154 member sample composed of the initial condition-based ensembles, the relationships between genesis and environmental precursors are investigated. We found that core water vapor from 700 to 850 hPa at 0000 UTC 24th had a large and positive impact on genesis 18 h later. Vertical shear, as expected, had a notable impact on genesis statistics, and the effect was strongest for vertical shear at 0300 UTC. After this time, shear as produced by Ernesto complicates the picture and weakens the relationship. It was found that low-level vorticity had a notable, positive relationship with genesis, while the relationship with mid-level vorticity was mixed. When examining genesis and nongenesis composites of the 850 hPa easterly wave at 0000 UTC, it was found that genesis was very positively correlated with the strength of this wave. For each of these relationships, the statistical significance was calculated over small ensemble samples and the corresponding p-values were found to often be below 5%. In other words, the genesis statistics associated with these 11 composited ensembles were found to be quite dependent on initial levels of local vorticity, vertical shear, and water vapor. These relationships were much stronger than was expected considering the high water vapor values present in the summer Tropics and the number of nondeveloping easterly waves.

When examining the SKEBS ensembles, we found that these showed notably

higher low-level θ bias compared with the other ensembles, as was noted by Charron et al. (2010). Vapor bias for the SKEBS ensembles, however, was notably *reduced* compared with the CNTL and breeding ensembles. The backscatter ensembles did effectively lessen 500 hPa geopotential height underdispersion, compared with our previous ensembles, in our forecasts of Ernesto, and thus demonstrated the effectiveness of the scheme in accounting for model error, which was not accounted for in our breeding and CNTL simulations. Near the TCs, all ensembles showed improved dispersion characteristics relative to the error, and compared to domain averages, yet SKEBS still showed higher spread and lower error in these regions than did the other ensembles. Compared with initial condition-based ensembles, it thus appears that the SKEBS scheme is better able to translate baroclinic and convective related error growth into ensemble spread in these forecasts of Hurricane Ernesto. While SKEBS did produce notably higher intra-ensemble TC positioning variability compared with the CNTL, the latter produced comparatively higher *intensity* variability.

We set up a Typhoon Nuri (2008) forecast case to better determine the durability of our ensemble results by using several of the same configurations for ensemble forecasts of both this and Ernesto. Despite the fact that Nuri was based in the Pacific, we find that the FNLq3 breeding forecasts of Nuri show notable similarities in terms of spatial perturbation patterns and spread, compared with forecasts of Hurricane Ernesto. Results from the FNLq3_NURI ensemble confirm that perturbations of vapor by themselves are not sufficient to create realistic ensembles, despite the importance of vapor to TC development. While a spatially-invariant scaling factor worked well for Ernesto, our FNLuvqt3_NURI ensemble produces spurious upper-level θ error growth, which leads to convergence among the members in terms of the lower levels over time. Attempts to rectify this with four-level vertical localization did not solve the problem as θ

error growth was too high for so few vertical rescaling factors.

The stochastic backscatter-based simulation of Typhoon Nuri confirms that Charron et al. (2010) is correct to emphasize the fact that SKEBS produces notably high θ bias in the low-levels, while, at the same time, producing remarkably *low* bias in terms of water vapor. In this same BS_CNTL_NURI ensemble, we find a remarkable amount of spread among the members in terms of TC intensity and spatial positioning compared with the other Nuri ensembles. At the same time, BS_CNTL_NURI is able to forecast TCs that are significantly more intense than those in the corresponding breeding schemes. While reasonable, the bias for both water vapor and θ are significantly higher in the Nuri versus the Ernesto case. Despite the fact that such differences are expected, their quantification does provide a realistic idea of the benefits and drawbacks associated with regional ensemble forecasts of TC genesis.

Overall, for the potential user of regional TC ensemble forecasts, we find that downscaling and creating a high-resolution GEFS-based ensemble provides TCs which show remarkable intensity variability, while also producing a higher-genesis likelihood than the large-scale GEFS. This latter point is especially of interest when the large-scale GEFS has difficulty forecasting a well-developed storm in any of its members. It is also found that the SKEBS scheme adds much-needed variance in terms of TC spatial positioning among the members, compared to downscaled GEFS, and it also reduces underdispersion across several metrics compared with other ensembles. These latter two perturbation methods should take precedence over multiphysics and breeding schemes, unless the latter can be done with a rescaling across the entire state vector while combining this with a vertical localization.

This study represents a comprehensive investigation of the error growth and dimensionality in tropical cyclone genesis environments at high resolution. Although

many results are obtained from the BGM scheme, the conclusions could be extended to similar applications with other ensemble methods. For example, the impacts of perturbing various variables on resulting TC genesis characteristics, the changes in error dimensionality by distance from the TC center, the benefits and drawbacks of downscaled GEFS versus multiphysics and SKEBS TC ensemble simulations, and the effects of TC precursors on genesis likelihood all have profitable use well beyond the limited scope of the BGM-related research.

Future work would perform similar calculations on forecasts of several more TC cases to better establish the significance of our general results. In terms of investigating the impact of environmental precursors on TC genesis, the robustness would be greatly enhanced by examining a sample which contains ensemble forecasts of several TCs both in the Atlantic and Pacific basins, as the genesis mechanisms appear to be slightly different in each. In terms of further evaluating the underdispersion of our ensembles, it would be beneficial to compare our BGM and SKEBS ensembles with forecasts created using Ensemble Kalman filter (EnKF) data assimilation. Such comparisons with EnKF, when possible, would also help us better determine the amount of error in the corresponding FNL analyses and evaluate the benefits provided by incorporating observations into the perturbation process.

REFERENCES

Aberson, S. D., 2001: The ensemble of tropical cyclone track forecasting models in the North Atlantic basin (1976–2000). *Bull. Amer. Meteor. Soc.*, **82,** 1895–1904.

Berner, J., M. L. G. Shutts, and T. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66,** 603–626.

----, S.-Y. Ha, J. P. Hacker, A. Fournier, and C. Snyder, 2011: Model uncertainty in a mesoscale ensemble prediction system: stochastic versus multiphysics representations. *Mon. Wea. Rev.*, **139,** 1972-1995.

Bishop, C. H., and Z. Toth, 1999: Ensemble transformation and adaptive observations. *J. Atmos. Sci.*, **56**, 1748-1765.

----, C. H., B J. Etherton, and S. Majumdar, 2001. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420–436.

Bister, M., and K. A. Emanuel, 1997: The Genesis of Hurricane Suillermo: TEXMEX analyses and a modeling study. *Mon. Wea. Rev.*, **125,** 2662-2682.

Blumen, W., 1978: Uniform potential vorticity flow: Part I. Theory of wave interactions and two-dimensional turbulence. *J. Atmos. Sci.*, **35,** 774–783.

Bowler, N. E., A. Arribas, S. E. Beare, K. R. Mylne, and G. J. Shutts, 2009: The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **134,** 703–722.

Buckingham, C., T. Marchok, I. Ginis, L. Rothstein, D. Rowe, 2010: Short- and medium-range prediction of tropical and transitioning cyclone tracks within the NCEP Global Forecasting System. *Wea. Forecasting.*, **25,** 1736-1754.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125,** 2887-2908.

Charron, M., G. Pellerin, L. Spacek, P. L. Houtekamer, N. Gagnon, H. L. Mitchell, and L. Michelin, 2010: Toward random sampling of model error in the Canadian Ensemble Prediction System. *Mon. Wea. Rev.*, **138,** 1877–1901.

Chen, S.-H., and W.-Y. Sun, 2002: A one-dimensional time dependent cloud model. *J. Meteor. Soc. Japan*, **80**, 99-118.

Chen, S. S, and W. M. Frank, 1993: A numerical study of the genesis of extratropical convective mesovortices. Part I: Evolution and dynamics. *J. Atmos. Sci.*, **50**, 2401-2426.

Cheung, K. K. W., and R. L. Elsberry, 2002: Tropical cyclone formations over the western north Pacific in the Navy Operational Global Atmospheric Prediction System forecasts. *Wea. Forecasting*, **17**, 800-820.

Davis, C. A., and L. F. Bosart, 2001: Numerical simulations of the genesis of Hurricane Diana (1984). Part I: Control simulation. *Mon. Wea. Rev.*, **129**, 1859-1881.

DeMaria, M., J. A. Knaff, and B. H. Connell, 2001: A tropical cyclone genesis parameter for the Tropical Atlantic. *Wea. Forecasting*, **16**, 219-233.

Descamps, L., and O. Talagrand, 2007: On some aspects of the definition of initial conditions for ensemble prediction. *Mon. Wea. Rev.*, **135**, 3260-3272.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107.

Dunkerton, T. J., M. T. Montgomery, and Z. Wang, 2009: Tropical cyclogenesis in a tropical wave critical layer: easterly waves. *Atmos. Chem. Phys.*, **9**, 5587-5646.

Epstein, E. S., 1969: The role of initial uncertainties in prediction. *J. Appl. Meteor.*, **8**, 190-198.

----, 1969: Stochastic dynamic prediction. *Tellus*, **6**, 739-759.

Frank, W. M., and E. A., Ritchie, 2001: Effects of vertical wind shear on the intensity and structure of numerically simulated hurricanes. *Mon. Wea. Rev.*, **129**, 2249-2269.

Fritsch, J. M., J. Hilliker, J. Ross, and R. L. Vislocky, 2000. Model consensus. *Wea. Forecasting* **15**, 571–582.

Gray, W. M., 1968: Global view of the origin of tropical disturbances and storms. *Mon. Wea. Rev.*, **96**, 969-700.

Grell, G. A. and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geoph. Res. Let.*, **29**, NO 14., 10.1029/2002GL015311, 2002.

Hagedorn, R., F. J. Doblas-Reyes, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus*, **57A**, 219–233.

Harr, P. A., R. L. Elsberry, and J. C. L. Chan, 1996: Transformation of a large monsoon depression to a tropical storm during TCM-93, *Mon. Wea. Rev.*, **124**, 2625–2643, 1996

Hendricks, E. A., M. T. Montgomery, and C. A. Davis, 2004: The role of "vortical" hot towers in the formation of Tropical Cyclone Diana (1984). *J. of Atmos. Sci.*, **61**, 1209-1232.

Hennon, C. C., and J. S. Hobgood, 2003: Forecasting tropical cyclogenesis over the Atlantic basin using large-scale data. *Mon. Wea. Rev.*, **131**, 2927–2940.

Hohenegger, C. and C. Schar, 2007: Atmospheric predictability at synoptic versus cloud-resolving scales. *Bull. Amer. Meteor. Soc.*, **88**, 1783–1793.

Hong, S. Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341.

Houtekamer, P. L., L. Lefaivre., J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225-1242.

Kalnay, E., 2006: Atmospheric modeling, data assimilation, and predictability. Cambridge University Press, 341pp.

Kerns, B., K. Greene, and E. Zipser, 2008: Four years of tropical ERA-40 vorticity maxima tracks. Part I: Climatology and vertical vorticity structure. *Mon. Wea. Rev.*, **136**, 4301–4319.

Kharin, V. V., and F. W. Zwiers, 2002: Climate predictions with multimodel Ensembles. *J. Climate*, **15**, 793–799.

Knaff, J. A., S. A. Seseske, M. DeMaria, and J. L. Demuth, 2004: On the influences of vertical wind shear on symmetric tropical cyclone structure derived from AMSU. *Mon. Wea. Rev.*, **132**, 2503-2510.

Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal cli- mate. *J. Climate*, **13**, 4196–4216.

Kuhl, D., and Coauthors, 2007: Assessing predictability with a local ensemble Kalman filter. *J. Atmos. Sci.*, **64**, 1116–1140.

Leith, C.E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409-418.

Li, X., M. Charron, L. Spacek, and G. Candille, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462.

Lin, J. W.-B., and J. D. Neelin, 2002: Considerations for stochastic convective parameterization. *J. Atmos. Sci.*, **59**, 959–975.

Liu, H., J. Anderson, Y.-H. Kuo, 2012: Improved analyses and forecasts of Hurricane Ernesto's genesis using radio occultation data in an ensemble filter assimilation system. *Mon. Wea. Rev.*, **140**, 151-166.

Lorenz, E. N., 1963: The predictability of hydrodynamic flow. *Transactions of the New York Academy of Sciences*, **4**, No 4, 409-432.

----, 1969: The predictability of a flow which possesses many scales of motion. *Tellus*, **21**, 289–307.

Mackey, B. P., and T. N. Krishnamurti, 2001: Ensemble forecast of a typhoon flood event. *Wea. Forecasting*, **16**, 399–415.

Mason, P. J., and D. J. Thomson, 1992: Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78.

McBride, J. L., and R. Zher, 1981: Observational analysis of tropical cyclone formation. Part II: Comparison of non-developing versus developing storms. *J. Atmos Sci.*, **38**, 1132-1151.

McLay, J. G., C. H. Bishop, and C. A. Reynolds, 2008: Evaluation of the ensemble transform analysis perturbation scheme at NRL. *Mon. Wea. Rev.*, **136**, 1093-1108.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmosphere: RRTM, a validated correlated-k model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682.

Montgomery, M. T., and J. Enagonio, 1998: Tropical cyclogenesis via convectively forced vortex Rossby waves in a three-dimensional quasigeostrophic model. *J. Atmos. Sci.*, **55**, 3176–3207.

----, M. E. Nicholls, T. A. Cram, and A. B. Saunders, 2006: A vortical hot tower route to tropical cyclogenesis. *J. Atmos. Sci.*, **63**, 355-386.

----, Z. Wang, and T. J. Dunkerton, 2010: Coarse, intermediate and high resolution numerical simulations of the transition of a tropical wave critical layer to a tropical storm. *Atmos. Chem and Phys.*, **10**, 10803-10827.

Nastrom, G. D., and K. S. Gage, 1985: A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **42**, 950–960.

Nolan, D. S., E. D. Rappin, and K. A. Emanuel, 2007: Tropical cyclogenesis sensitivity to environmental parameters in radiative-convective equilibrium. *Q. J. R. Met. Soc.*, **133**, 2085-2107.

Oczkowski, M., I. Szunyogh, and D. J. Patil, 2005: Mechanisms for the development of locally low-dimensional atmospheric dynamics. *J. Atmos. Sci.*, **62**, 1135–1156.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56**, 415–428.

Palmer, T. N., F. Molteni, R. Mureau, and R. Buizza, 1993: Ensemble prediction. *Proceedings of the ECMWF seminar on validation of models over Europe:* 1. ECMWF, Shinfield Park, Reading, UK, 21-66.

----, 1997: On parameterizing scales that are only somewhat smaller than the smalles reslved scales with application to convection and orography. *Proceedings of the ECMWF workshop on new insights and approaches to convective parameterization*, 4-7 November 1996. ECMWF, Shinfield Park, Reading, UK, 328-337.

----, 2003: Representation of model uncertainty in en-semble prediction systems. *Proceedings of the Ninth ECMWF Workshop on Meteorological Operational Systems.* ECMWF, Shinfield Park, Reading, United Kingdom, 46–53.

Patil, D. J., B. R. Hunt, E. Kalnay, J. A. Yorke, and E. Ott, 2001: Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878–5881.

Reasor P. D., Montgomery M. T., Bosart L. F. 2005. Mesoscale observations of the genesis of Hurricane Dolly (1996). *J. Atmos. Sci.*, **62**, 3151 – 3171.

Reynolds, C. A., J. Teixeira, and J. G. McLay, 2008: Impact of stochastic convection on the ensemble transform. *Mon. Wea. Rev.*, **136**, 4517–4526.

Ritchie, E. A., 1995: Mesoscale aspects of tropical cyclone formation. Ph.D. dissertation, Centre for Dynamical Meteorology and Oceanography, 167 pp. [Available from Centre for Dynamical Meteorology and Oceanography, Applied Mathematics, Monash University, Melbourne, VIC 3168, Australia.]

-----, and G. J. Holland, 1997: Scale interactions during the formation of Typhoon Irving. *Mon. Wea. Rev.*, **125**, 1377– 1396.

Rogers, R., S. Aberson, M. Black, P. Black, J. Cione, P. Dodge, J. Dunion, J. Gamache, J. Kaplan, M. Powell, N. Shay, N. Surgi, and E. Uhlhorn, 2006: The Intensity Forecasting Experiment: A NOAA multiyear field program for improving tropical cyclone intensity forecasts. *B. Am. Meteorol. Soc.*, **87**, 1523–1537.

Rotunno, R., and C. Snyder, 2008: A generalization of Lorenz's model for the predictability of flows with many scales of model for the predictability of flows with many scales of motion. J. Atmos. Sci., **65**, 1063–1076.

Shutts, G. J., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.

Simpson, J., E. Ritchie, G. J. Holland, J. Halverson, and S. Stewart, 1997: Mesoscale interactions in tropical cyclone genesis. *Mon. Wea. Rev.*, **125**, 2643–2661.

Sippel, J. A., and F. Zhang, 2008: A probabilistic analysis of the dynamics and predictability of tropical cyclogenesis. *J. Atmos Sci.*, **65**, 3440-3459.

Skammarock, W.C., J.B. Klemp, J. Dudhia, D.O. Gill, D.M. Barker, M.G. Duda, X.-Y. Huang, W. Wang, and J.G. Powers, 2008: A description of the Advanced Research WRF Version 3. NCAR Tech. Note NCAR/TN-475+STR., June 2008, 133 pp.

Slade, S., and E. Maloney, 2012: An Intraseasonal Prediction Model of Atlantic and East Pacific Tropical Cyclone Genesis. *Mon. Wea. Rev.* doi:10.1175/MWRD-12-00268.1, in press.

Snyder, A., Z. Pu, and Y. Zhu, 2010: Tracking and verification of the east Atlantic tropical cyclone genesis in NCEP global ensemble: Case studies during NASA African monsoon multi-disciplinary analyses. *Wea. Forecasting*, **25**, 1397–1411.

----, Z. Pu, and C.A. Reynolds, 2011: Impact of stochastic convection on ensemble forecasts of tropical cyclone development. *Mon. Wea. Rev.*, **139**, 620-626.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the NCEP global model. *Tellus*, **57A**, 528–545.

Teixeira, L., and C. A. Reynolds, 2008: Stochastic nature of physical parameterizations in ensemble prediction: A stochastic convection approach. *Mon. Wea. Rev.*, **136**, 483–496.

Tennant, W. J., G. J. Shutts, A. Arribas, and S. A. Thompson, 2011: Using a stochastic

kinetic energy backscatter scheme to improve MOGREPS probabilistic forecast skill. *Mon. Wea. Rev.*, **139**, 1190–1206.

Torn, R. D., G. J. Hakim, and C. Snyder, 2006: Boundary conditions for limited-area ensemble Kalman filters. *Mon. Wea. Rev.*, **134**, 2490-2502.

----, ----, 2008: Ensemble-based sensitivity analysis. *Mon. Wea. Rev.*, **136**, 663-677.

----, 2010: Performance of a mesoscale Ensemble Kalman Filter (EnKF) during the NOAA high-resolution hurricane test. *Mon. Wea. Rev.*, **138**, 4375-4392.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317– 2330.

----, and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.

Wang, X. and Bishop, C. H. 2003. A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.

Wei, M., Z. Toth, R. Wobus, Y. Zhu, and C. H. Bishop, 2004: Initial perturbations for the NCEP ensemble forecast system. *Proc. First THORPEX Int. Science Symp.*, Montreal, QC, Canada, World Meteorological Organization, 227–230.

----, ----, ----, and ----, 2006: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global ensemble forecast system. NOAA/NCEP Office Note 453, 33 pp.

----, ----, ----, and ----, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus*, **60A**, 62–79.

Zhang, Z., and T. N. Krishnamurti, 1997: Ensemble forecasting of hurricane tracks., *Bull. Amer. Meteor. Soc.*, **78**, 2785-2795.